An Intellectual Property Data Access Control Method for Crowdsourced Testing System

Song Huang  
Army Engineering University of PLA  
Nanjing, Jiangsu, China  
hs0317@163.com

Changyou Zheng  
Army Engineering University of PLA  
Nanjing, Jiangsu, China  
zheng_chy@163.com

Zhen Yang  
Army Engineering University of PLA  
Nanjing, Jiangsu, China  
yangzhenlgd@163.com

Jinyong Wan  
Army Engineering University of PLA  
Nanjing, Jiangsu, China  
jy_wan_wy@163.com

Abstract—In the crowdsourced testing system, due to the openness of crowdsourced testing platform and other factors, the security of crowdsourced testing intellectual property cannot be effectively protected. We proposed an attribute-based double encryption scheme, combined with the blockchain technology, to achieve the data access control method of the code to be tested. It can meet the privacy protection and traceability of specific intellectual property in the crowdsourced testing environment. Through the experimental verification, the access control method is feasible, and the performance test is good, which can meet the normal business requirements.

Keywords: crowdsourced testing; data access control; intellectual property; attribute-based encryption; blockchain

I. INTRODUCTION

Due to the rapid change of software, the cycle of software testing is drastically reduced. Therefore, how to get a lot of real test feedback quickly and complete the test task at a lower cost is one of the difficulties in the field of software testing. Crowdsourcing is a distributed problem solving and production organization mode brought by the Internet, which can effectively solve this problem in the field of software testing [1]. Crowdsourcing technology is applied to software testing, and a large number of online workers participate in the testing task. It can provide a good simulation of real application scenarios and users, with short test cycle and low cost [2][3]. Therefore, crowdsourced testing technology has been widely concerned by academia and industry.

As an intelligent system, crowdsourced testing system is facing the problem of data security. Due to the openness of the system, anyone can view the test tasks. The intellectual property of the crowdsourced testing cannot be effectively protected. The task requesters are afraid to publish the code of the tested software in clear text on the crowdsourced testing platform. This greatly dampens the initiative of task requesters to submit test requirements on the platform, which is not conducive to the long-term development of crowdsourced testing. The intellectual property security of crowdsourced testing system needs to be solved.

The common way to protect the security and privacy of data is to encrypt the data. The intellectual property is encrypted first. After the crowdsourced testing workers receive the ciphertext of the intellectual property, decrypt the ciphertext. The common encryption methods are symmetric encryption algorithm and asymmetric cryptographic algorithm. Symmetric encryption algorithm has great efficiency, but it is difficult to manage key. Asymmetric cryptographic algorithm is more convenient in data sharing, but the encryption speed is slow. In crowdsourced testing activities, the data needs to be distributed to multiple crowdsourced testing workers. If traditional asymmetric encryption is used, it takes a lot of time to encrypt intellectual property data to different workers. In this paper, attribute-based encryption (ABE) is used. The task requester formulates the access policy, which can specify the crowdsourced testing workers with certain attributes to access the corresponding intellectual property data. The crowdsourced testing intellectual property is a large-scale data. In order to improve the encryption speed, we use the symmetric encryption algorithm to encrypt the metadata, and then use the ABE algorithm to encrypt the symmetric key. In the choice of symmetric encryption algorithm, we use Advanced Encryption Standard (AES), which is the most popular symmetric encryption algorithm. The ciphertext data is stored in the Interplanetary file system (IPFS), the ciphertext of hash address and key is stored on the blockchain to realize the traceability function.

The contributions of this paper are as follows:

- An intellectual property data access control method for crowdsourced testing system is proposed. The symmetric encryption algorithm and attribute-based encryption method are combined to realize double encryption.
- Combine blockchain technology with IPFS to realize trusted storage of intellectual property data. According to the non-tampering characteristics of blockchain, the traceability of intellectual property data can be realized.
II. RELATED WORK

A. Attribute-based Encryption

Attribute-based encryption technology is a new encryption technology, which was first proposed by Waters et al [4]. It realized one-to-many encryption and decryption. Attribute-based encryption can be divided into two categories: attribute encryption of ciphertext policy (CP-ABE) and attribute encryption of key policy (KP-ABE). In CP-ABE, ciphertext is associated with access policy, and decryption key is associated with attribute. Only when the attribute in the key satisfies the access policy can it be decrypted. While in KP-ABE, ciphertext is associated with attribute, and decryption key is associated with access policy [5]. Attribute-based encryption technology can achieve fine-grained data access control.

Since CP-ABE was proposed, scholars have done a lot of research in the field of access control. Lai et al. proposed a new model for CP-ABE with partially hidden access structures [6]. Helil et al. introduced the concept of sensitive data set constraint, and proposed a CP-ABE access control scheme with sensitive data set constraint hidden attribute [7]. Bramm et al. applied CP-ABE to distributed data sharing and access protocol based on blockchain, which improved the security of key management in distributed system [8]. Banerjee et al. present a secure fine-grained user access control scheme for data usage in the IOT environment. The scheme supports multi-authority ABE [9].

B. Data Privacy Protection in Crowdsourcing

With the development of crowdsourcing technology, more and more scholars pay attention to privacy protection in crowdsourcing. The traditional way of protecting data privacy is based on encryption. Sara Foresti, an Italian scholar, proposed a method of combining data fragmentation and encryption to protect data privacy [10]. In the process of crowdsourcing, data can be published or stored on an untrusted server to achieve distributed access with lower consumption and higher availability. Sun et al. proposed using blockchain to solve the privacy problem of crowdsourcing code. Use the local code granulation tool to divide and encrypt the code to prevent the code from being completely leaked [11]. In crowdsourced testing system, crowdsourced testing workers need to obtain all the code to complete the test. Therefore, the above code segmentation method cannot be directly used in crowdsourced testing. Cui et al. proposed a privacy protection framework based on blockchain, which uses homomorphic encryption algorithm to ensure the privacy and integrity of crowdsourcing data [12]. But it only considers the malicious allocation of crowdsourcing rewards, the privacy protection of crowdsourcing tasks is not considered.

In the crowdsourced testing system, there is no effective scheme to protect the privacy of the code to be tested, which can support the effective testing of the workers and meet the personalized privacy protection needs of different task requesters. Therefore, we propose an attribute-based encryption scheme, combined with blockchain technology, to achieve the data access control method of the code to be tested.

This method can satisfy the privacy protection of specific intellectual property in the crowdsourced testing system.

III. SYSTEM MODEL AND METHOD

A. System Model

We implement an intellectual property data access control method for crowdsourcing testing system. Among them, the crowdsourced testing intellectual property refers to the code to be tested submitted by the task requester. This method mainly includes five types of entities, which are crowdsourced testing task requester, multiple crowdsourced testing workers, crowdsourcing testing platform [13], blockchain (alliance chain) [14] and IPFS [15].

- **Crowdsourced testing task requester**: the owner of crowdsourced testing intellectual property data, who submits the test requirements and software code to the crowdsourcing testing platform. The task requester specifies the access control policy and sends it to the crowdsourcing testing platform.
- **Crowdsourced testing workers**: the person who completes the crowdsourced testing tasks. They can access the intellectual property data of crowdsourced testing according to the access policy.
- **Crowdsourced testing platform**: the platform which provide online systems for task requesters and workers. The platform is responsible for generating keys, setting parameters, registering users, distributing test tasks and integrating test reports. It is reliable to the workers and task requesters.
- **Blockchain (alliance chain)**: the alliance chain node is composed of crowdsourced testing platform, task requester and workers. Multi party joint maintenance alliance chain. Alliance chain stores data identification number, hash address of intellectual property ciphertext, AES key encrypted by attribute-based encryption and other information.
- **IPFS**: a distributed file system based on content addressing, which provides storage space under the chain and stores intellectual property ciphertext information.

B. Method

The intellectual property data access control method for crowdsourced testing system can be divided into the following steps, and shown in Figure 1:

- **Initialize.** The crowdsourced testing platform executes the initialization algorithm $\text{Setup}(\lambda)$, input the security parameters $\lambda$. The public parameter $pk$ and master key $mk$ are generated.
- **Registration and private key distribution.** The crowdsourced testing workers submit personal registration information to the platform, which includes the workers’ occupation, age, hobbies and other information. The platform verifies the identity of workers and assigns an attribute set $s$ according to the characteristics of workers. The platform executes key distribution algorithm $\text{KeyGen}(mk, s)$, inputs
the master key $mk$ and attribute set $s$ to get the worker's private key $sk$. The platform distributes the private key $sk$ to the workers.

- **The first encryption.** The AES key $K_A$ is generated randomly, and the requester uses AES symmetric encryption algorithm to encrypt the intellectual property metadata $M$ to get the intellectual property ciphertext $C_M = \text{Encrypt}_{AES} (K_A, M)$.

- **Ciphertext stored off-chain.** The requester uploads the ciphertext $C_M$ to IPFS, and then IPFS returns a unique hash address $Address (C_M)$.

- **The second encryption.** According to the characteristics of the task to be tested, the requester formulates the access control strategy $T$ and submits it to the platform. The platform executes the ciphertext generation algorithm, inputs common parameters $pk$, access control policy $T$, AES key $K_A$, and get the ciphertext $C_{K_AT} = \text{Encrypt}_{CP-ABE} (pk, K_A, T)$.

- **Upload to blockchain.** The crowdsourced testing platform uploads the data ID, ciphertext hash address $Address (C_M)$ and ciphertext $C_{K_AT}$ to the alliance chain.

- **Query on blockchain.** Combined with the access control policy submitted by requester and the attributes of the authorized workers, the platform distributes the data ID to workers who meet access conditions. According to data ID, the workers query in the alliance chain, get the hash address of intellectual property ciphertext $Address (C_M)$ and the encrypted symmetric key $C_{K_AT}$.

- **The first decryption.** According to the private key $sk$, the worker decrypts the encrypted symmetric key $C_{K_AT}$. If the worker satisfies the attribute set required in the access control policy, the plaintext symmetric key can be obtained according to the decryption algorithm $K_A = \text{Decrypt}_{CP-ABE} (C_{K_AT}, sk, pk)$.

- **The second decryption.** According to the hash address of intellectual property ciphertext $Address (C_M)$, the workers obtain intellectual property ciphertext in IPFS. Then the symmetric key $K_A$ is used to decrypt it, the plaintext information of the intellectual property metadata can be obtained $M = \text{Decrypt}_{AES} (K_A, C_M)$.

In the above process, double encryption method is adopted. Because of the slow speed of attribute-based encryption, if the metadata is encrypted directly by attribute-based encryption algorithm, it will get lower efficiency. In order to improve the efficiency of the algorithm, we choose to use symmetric encryption algorithm AES to encrypt intellectual property metadata. Then AES key is encrypted by attribute-based encryption method, so as to realize one-to-many fine-grained access control. Only workers who meet the access policy can decrypt it. We use double encryption to combine the advantages of ABE and symmetric encryption. It not only realizes fast encryption, but also realizes fine-grained access control.

As for data storage, due to the limited storage space of blockchain, each node has to store information. Considering the persistence of blockchain data, if large-scale data is stored on the blockchain, it will inevitably lead to data expansion and system performance degradation. Therefore, intellectual property ciphertext data is stored off-chain, IPFS is used as the database off-chain. And store hash address et al. on the chain.

By storing the data related to intellectual property and the information of workers who downloaded the data into the blockchain, the traceability of intellectual property access is ensured.
process can be realized. The platform uploads the hash address and key of intellectual property, invokes the smart contract, and initiates the information storage transaction, so as to store the data related to intellectual property on the alliance chain. When the workers access the data on the alliance chain, they initiate a transaction and store the worker's identity information and the accessed data information on the alliance chain, so as to realize the traceability of the access process.

IV. EXPERIMENTAL ANALYSIS

In order to verify the feasibility of the intellectual property data access control method for crowdsourced testing system, we construct a prototype system to verify. The hardware environment of the experiment is ecs.c5.large, 2-core 4G, and the operation system is Ubuntu16.04. We built an alliance chain with Hyperledger Fabric1.4.0 [16]. Java is chosen as the main programming language to implement encryption algorithm. The interaction with IPFS and Hyperledger Fabric is realized by integrating fabric-sdk-java, ipfs-api and Docker.

We test the performance of the attribute-based encryption algorithm. Specifically, test the influence of attributes on the encryption and decryption time. The experimental results are shown in Figure 2. The experimental results show that with the increase of attributes, the encryption and decryption time of symmetric key increases almost linearly. The time consumption is short, and the performance is good.

![Figure 2. Encryption and decryption time under different number of attributes](image)

We also test the performance of the blockchain, including the Invoke method and the Query method of the smart contract. The Invoke method corresponds to the upload operation of intellectual property, and the Query method corresponds to the query operation. Under different concurrency, test the time cost. The results of the test are shown in Figure 3. When the concurrency increases gradually, the time cost of Invoke method and Query method also increases gradually, but it is still in a reasonable range. In the case of high concurrency, blockchain can still work normally, and its performance meets the requirements of high concurrency.

We compare different encryption systems, and the results are shown in Table 1. The results show that the system we proposed performs better in the aspects of access control, data security, traceability and storage performance.

![Table 1. The comparison between the encryption systems](image)

V. CONCLUSION

In order to solve the security and privacy problems of intellectual property in the crowdsourced testing system, we propose an intellectual property data access control method for crowdsourced testing system. Through double encryption, the task requester can customize the access policy to meet the personalized privacy protection needs. Combined with blockchain technology and IPFS, the trusted storage and traceability of intellectual property are realized. The experimental results show that the scheme is feasible and the performance test is good. But there are still some imperfections in this method, such as flexible revocation of attributes, dynamic update of access policies, etc., which will be our next research direction.

ACKNOWLEDGMENT

The project is supported by National Key R&D Program of China (No: 2018YFB1403400), Natural Science Foundation of China (No: 61702544).

REFERENCES


