SkyEye: A Traceable Scheme for Blockchain

Tianjun Ma
Haixia Xu*
matianjun@iie.ac.cn
xuhaixia@iie.ac.cn
State Key Laboratory of Information Security, Institute of Information Engineering, CAS, Beijing, China
School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
Data Assurance and Communication Security Research Center, Chinese Academy of Sciences, Beijing, China

Peili Li
lipeili@iie.ac.cn
State Key Laboratory of Information Security, Institute of Information Engineering, CAS, Beijing, China
Data Assurance and Communication Security Research Center, Chinese Academy of Sciences, Beijing, China

ABSTRACT

Many studies focus on the blockchain privacy protection. Unfortunately, the privacy protection brings regulatory issues (e.g., countering money-laundering). Tracing users' identities is a critical step in addressing blockchain regulatory issues. In this paper, we propose SkyEye, a traceable scheme for blockchain. SkyEye can be applied to the blockchain applications that satisfy the following conditions: (I) The users have public and private information, where the public information is generated by the private information; (II) The users' public information is disclosed in the blockchain data. SkyEye enables the regulator to trace users' identities. The design of SkyEye leverages some cryptographic primitives, including chameleon hash and zero-knowledge succinct non-interactive arguments of knowledge (zk-SNARKs). Moreover, we demonstrate the security of SkyEye under specific cryptographic assumptions. Finally, we implement two prototypes of SkyEye, and evaluate the running time and related data storage requirements by performing the aforementioned prototypes.

KEYWORDS

Traceable scheme, blockchain, chameleon hash, zero knowledge

1 INTRODUCTION

The blockchain was first introduced in Bitcoin[35], and quickly became the supporting technology of decentralized cryptocurrencies such as PPcoin[25] and Litecoin[1]. The blockchain integrates multiple technologies (e.g., cryptography and peer-to-peer networking) and includes a variety of features: distributed, decentralized, anonymity, transparency, and so on. Today, the blockchain is not only applied in decentralized cryptocurrencies, but also has broad applications in other fields, including defense, finance, and smart contract.

The blockchain can be considered a distributed database that only appends data (e.g., transactions). The data is stored in the block that contains the block header and block body. Every block header includes the hash of the previous block, forming a chain. The strategy of appending a block to the blockchain uses a consensus mechanism such as proof of work (PoW)[35], proof of stake (PoS) [5, 14, 24], or practical byzantine fault tolerance (PBFT)[10].

*The corresponding author.

In many blockchain applications, every user generally has public/private information (e.g., the public key address and the signature private key for each user in Bitcoin[35], more details about the public/private information are described in Section 2.3).

Many studies focus on the blockchain privacy protection [7, 38]. Unfortunately, the privacy protection brings regulatory issues. On one hand, if a user's private information is lost or stolen, the user loses control of the data corresponding to the private information forever. For example, if the Bitcoin's user loses the signature private keys in his wallet, there is no way to recover the coins in this wallet. In other words, the user loses the coins controlled by these signature private keys forever. On the other hand, strong privacy protection in the blockchain facilitates many criminal activities (e.g., ransomware[2], money laundering). CipherTrace's second quarter 2019 cryptocurrency anti-money laundering report shows that the total amount of funds that cybercriminals directly steal, scam, and misappropriate from users and trading platforms is approximately $4.4 billion in aggregate for 2019. These regulatory issues not only present a serious threat to the interests of users, but also have seriously hindered the development and application of the blockchain.

We stress that tracing users' identities is a critical step in addressing blockchain regulatory issues. When each user's identity in the blockchain data is determined, the regulator can conduct some regulatory operations (such as Big Data analysis) to decide who should be punished or who should own the lost data. Although there has been progress in designing traceable mechanisms, such as zkLedger[36] and several others[4, 15, 20, 23], these approaches are designed for specific application environment and do not seem to have been extended to other applications; see Section 9 for more details.

Our contributions. In this paper, we present a traceable scheme that can be applied to a class of blockchain applications. The main contributions of this paper are as follows.

First, we introduce the notion of a traceable scheme for blockchain and formalize the security properties to be satisfied, namely *identity proof indistinguishability and identity proof unforgeability*.

Second, we propose SkyEye, a traceable scheme for blockchain. SkyEye can be applied to the blockchain applications that satisfy the following conditions: (I) The users have public and private information, where the public information is generated by the private information; (II) The users' public information is disclosed

in the blockchain data. These blockchain applications are called **SkyEye-friendly blockchain applications**; see Section 2.3 for details. SkyEye requires the user to register only once, and enables the regulator to trace users' identities. In our design strategy, we add identity proofs, associated with the users' private information, to the blockchain data. SkyEye is designed by using some cryptographic primitives (including chameleon hash[27] and zk-SNARKs[21]). In addition, we demonstrate the security of SkyEye under specific cryptographic assumptions.

Finally, we implement two prototypes of SkyEye: $SkyEye_H$ and $SkyEye_S$. These correspond to the two primary ways of generating public and private information in the blockchain applications. The first way is through a pseudorandom function, and the second way is using elliptic curve scalar multiplication. We evaluate the running time and related data storage requirements by performing $SkyEye_H$ and $SkyEye_S$. Our evaluation results illustrate that using an i7 processor, a 16 GB RAM desktop machine, and a Merkle tree depth of 34, the time taken by a verifier to verify a user's identity proof is nearly 4.6 ms in the first way and less than 25 ms in the second way.

Paper organization. The remainder of this paper is organized as follows. Section 2 provides the background. Section 3 provides key ideas in SkyEye design and an overview of SkyEye. Section 4 defines the algorithm and security of the traceable scheme for blockchain. Section 5 details SkyEye. Section 6 describes our implementation and the evaluation results. We present the potential applications of SkyEye in Section 7. We discuss remaining issues of SkyEye and future work in Section 8. We discuss related work in Section 9 and summarize this paper in Section 10.

2 BACKGROUND

2.1 Cryptographic Preliminaries

The cryptographic building blocks in our construction include the following: chameleon hash scheme, zk-SNARKs, and public key encryption. Below, we informally describe these notions.

Chameleon hash scheme. Compared with the traditional hash scheme, the chameleon hash scheme has a special property: the user who knows the trapdoor can easily find collision. A chameleon hash scheme $Chash = (\mathcal{G}_{chash}, \mathcal{K}_{chash}, \mathcal{H}_{chash}, \mathcal{CF}_{chash})$ is described below:

- $\mathcal{G}_{chash}(\lambda) \rightarrow pp_{chash}$. Given a security parameter λ, $\mathcal{G}_{chash}$ returns the public parameters pp_{chash}.
- $\mathcal{K}_{chash}(pp_{chash}) \rightarrow (pk_{chash}, sk_{chash})$. Given the public parameters pp_{chash}, $\mathcal{K}_{chash}$ returns a pair of public/private keys (pk_{chash}, sk_{chash}), where sk_{chash} is also known as the trapdoor.
- $\mathcal{H}_{chash}(pk_{chash}, m, r) \rightarrow CH$. Given the public key pk_{chash}, a message m, and a random number r, $\mathcal{H}_{chash}$ returns a chameleon hash value CH about m.
- $\mathcal{CF}_{chash}(sk_{chash}, m, m', r) \rightarrow r'$. Given the trapdoor sk_{chash}, two messages m, m', and the random number r, $\mathcal{CF}_{chash}$ returns r' such that $\mathcal{H}_{chash}(pk_{chash}, m, r) = \mathcal{H}_{chash}(pk_{chash}, m', r')$.

A chameleon hash scheme satisfies three secure properties: (i) *collision resistance*; (ii) *trapdoor collision*; and (iii) *semantic security*. More details are available in [3, 27].

There is a relationship between the public key pk_{chash} and the trapdoor sk_{chash}, which we refer to as the generation relationship. As in [27], the public parameters $pp_{chash} = (p, q, g)$, where p, q are prime numbers such that $p = kq + 1$, and the order of g is q in $\mathbb{Z}_p^*$. The public key $pk_{chash} = h$ is computed as follows: $h = g^x \bmod p$, where $x \in \mathbb{Z}_q^*$ is the trapdoor. Let equation $pk_{chash} = chash\_gen(sk_{chash})$ describe this relation, where $chash\_gen(\cdot)$ denotes the generation algorithm between pk_{chash} and sk_{chash}.

Zero-knowledge succinct non-interactive arguments of knowledge. Let $\mathcal{R}_{AC} = \{(x, w) \in \mathbb{F}^n \times \mathbb{F}^h | AC(x, w) = 0^l\}$ be an NP relation, where $\mathbb{F}$ denotes a finite field, and $AC : \mathbb{F}^n \times \mathbb{F}^h \rightarrow \mathbb{F}^l$ denotes an $\mathbb{F}$-arithmetic circuit. The language for $\mathcal{R}_{AC}$ is $\mathcal{L}_{AC} = \{x \in \mathbb{F}^n | \exists w \in \mathbb{F}^h \ s.t. \ AC(x, w) = 0^l\}$. A zk-SNARK scheme $NIZK = (\mathcal{K}_{nizk}, \mathcal{P}_{nizk}, \mathcal{V}_{nizk})$ corresponds to the language $\mathcal{L}_{AC}$, which is described below:

- $\mathcal{K}_{nizk}(\lambda, AC) \rightarrow (pk, vk)$. Given a security parameter λ and an $\mathbb{F}$-arithmetic circuit AC, $\mathcal{K}_{nizk}$ returns a proving/verification key pair (pk, vk).
- $\mathcal{P}_{nizk}(pk, x, w) \rightarrow \pi$. Given the proving key pk, a statement x, and a witness w, $\mathcal{P}_{nizk}$ returns a proof π for a statement x using a witness w.
- $\mathcal{V}_{nizk}(vk, x, \pi) \rightarrow \{0, 1\}$. Given the verification key vk, the statement x, and the proof π, $\mathcal{V}_{nizk}$ returns 1 if verification succeeds, or 0 if verification fails.

A zk-SNARK scheme satisfies five secure properties: (i) *completeness*; (ii) *soundness*; (iii) *succinctness*; (iv) *proof of knowledge*; and (v) *perfectly zero knowledge*. More details are available in [7, 9, 21].

Public key encryption. A public key encryption scheme $Enc = (\mathcal{G}_{enc}, \mathcal{K}_{enc}, \mathcal{E}_{enc}, \mathcal{D}_{enc})$ is described below:

- $\mathcal{G}_{enc}(\lambda) \rightarrow pp_{enc}$. Given a security parameter λ, $\mathcal{G}_{enc}$ returns the public parameters pp_{enc}.
- $\mathcal{K}_{enc}(pp_{enc}) \rightarrow (pk_{enc}, sk_{enc})$. Given the public parameters pp_{enc}, $\mathcal{K}_{enc}$ returns a pair of public/private keys (pk_{enc}, sk_{enc}).
- $\mathcal{E}_{enc}(pk_{enc}, m) \rightarrow c$. Given the public key pk_{enc} and a message m, $\mathcal{E}_{enc}$ returns a ciphertext c.
- $\mathcal{D}_{enc}(sk_{enc}, c) \rightarrow m/\perp$. Given the private key sk_{enc} and the ciphertext c, $\mathcal{D}_{enc}$ returns a message m, or returns $\perp$ if decryption fails.

The public encryption scheme Enc satisfies a security property: ciphertext indistinguishability under adaptive chosen ciphertext attack (IND-CCA2 security). More details are provided in [11].

2.2 Notation

We use u to denote a user, id_u to denote the identity of u, and $proof_{id_u}$ to denote u's identity proof. Let $(pk_{chash_u}, sk_{chash_u})$ denote u's chameleon hash public/private key pair and CH_{id_u} denote the chameleon hash value of identity id_u. We denote u's public/private information as $(pub_u, priv_u)$.

We use $pk_{chash_u}||CH_{id_u}$ to denote the concatenation of pk_{chash_u} and CH_{id_u}, where $||$ denotes the concatenate symbol. Let $MT = (rt; pk_{chash_1}||CH_{id_1}, ..., pk_{chash_n}||CH_{id_n})$ denote a Merkle tree, where rt denotes the root of the Merkle Tree and $(pk_{chash_i}||CH_{id_i})$ denotes one leaf node in the Merkle tree. Let (pk_{reg}, sk_{reg}) denote the encryption public/private key pair of the regulator. Moreover, we use $\boldsymbol{B_s}$ to denote SkyEye-friendly blockchain applications and $\boldsymbol{B_{se}}$ to denote $\boldsymbol{B_s}$ using SkyEye.

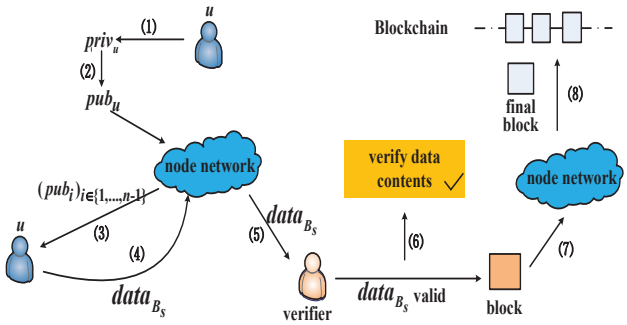

Figure 1: Overview of B_s.

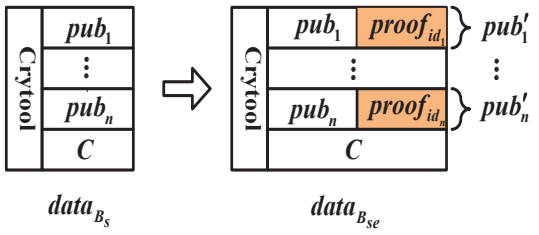

Figure 2: Design idea.

2.3 SkyEye-friendly blockchain applications

Here, we describe the blockchain data in $\boldsymbol{B_s}$ and an overview of $\boldsymbol{B_s}$.

2.3.1 The Blockchain Data in $\boldsymbol{B_s}$. $\boldsymbol{B_s}$ satisfies two conditions: (I) The users have public and private information, where the public information is generated by the private information; (II) The users' public information is disclosed in the blockchain data.

We use equation $pub = gen(priv)$ to describe the generation relation in the condition (I), where pub denotes the public information, $priv$ denotes the private information, and $gen(\cdot)$ denotes the generation algorithm between pub and $priv$, which has one-wayness, i.e., it is easy to compute pub using the private information $priv$ but is hard to invert. In many blockchain applications, every user generally has private information that corresponds to public information. For example, the public key address and the signature private key in Bitcoin[35] are the user's public/private information. In Zerocash[7], $(sn, (a_{sk}, \rho))$ is the user's public/private information, where sn is the serial number, a_{sk} is the address private key, and ρ is the random number used to generate the serial number. The public information is generated by the private information via a cryptographic method, such as the pseudorandom function, or elliptic curve scalar multiplication.

According to the condition (II), the blockchain data in $\boldsymbol{B_s}$ can be divided into two parts: one part is the users' public information, such as the input/output addresses in Bitcoin[35], and the other part is the data contents, such as the payment amount and the executable contract code. Therefore, the blockchain data in $\boldsymbol{B_s}$ can be represented as the equation $\boldsymbol{data_{B_s}} = [(pub_i)_{i\in\{1,\ldots,n\}}, C]_{crytool}$, where $(pub_i)_{i\in\{1,\ldots,n\}}$ denotes the set of the users' public information, n is the number of the users' public information in the blockchain data such as the number of pubic key addresses in a Bitcoin transaction, C denotes the data contents, and $crytool$ denotes the cryptographic tools (e.g., digital signature) that guarantee blockchain features such as tamper-resistance and privacy protection.

For example, Bitcoin[35], Ethereum[41], and RSCoin[13] are the applications that belong to $\boldsymbol{B_s}$. In these blockchain applications, the public key address and the signature private key are the user's public/private information, where the public key address is generated by the signature private key. Moreover, the user's public key address is disclosed in the blockchain data. We briefly describe how to use SkyEye in these three applications in Section 7.

2.3.2 An overview of $\boldsymbol{B_s}$. As shown in Figure 1, in (1) and (2), the user u generates $(pub_u, priv_u)$, and publishes pub_u to the node network. In (3), for creating data, the user u obtains others' public information $(pub_i)_{i\in\{1,\ldots,n-1\}}$ from the node network. In (4), the user u creates $\boldsymbol{data_{B_s}} = [(pub_i)_{i\in\{1,\ldots n-1,n\}}, C]_{crytool}$, where pub_n denotes the public information pub_u, and publishes $\boldsymbol{data_{B_s}}$ to the node network. In (5) and (6), a verifier receives $\boldsymbol{data_{B_s}}$ from the node network and verifies data contents. If the verification is successful, $\boldsymbol{data_{B_s}}$ is valid and is added to the block generated by the verifier. In (7), the block is published in the node network by the verifier. In (8), according to a consensus mechanism, the nodes in the network select a final block and add it to the blockchain.

3 KEY IDEAS AND SKYEYE OVERVIEW

In this section, we provide key ideas in SkyEye design and an overview of SkyEye.

3.1 Key Ideas

As shown in Figure 2, the design idea of SkyEye is to add identity proofs to $\boldsymbol{data_{B_s}}$. The blockchain data in $\boldsymbol{B_{se}}$ can be represented as the equation $\boldsymbol{data_{B_{se}}} = [(pub_i, proof_{id_i})_{i\in\{1,\ldots,n\}}, C]_{crytool}$, where $proof_{id_i}$ denotes the identity proof of the user whose identity is id_i, and the other variables are the same as those in the equation $\boldsymbol{data_{B_s}}$. $(pub_i, proof_{id_i})$ can be viewed as the new public information pub_i' of the user whose identity is id_i.

The identity proof is the core of SkyEye. The two purposes of the identity proof are to prove the user's legitimacy and to achieve tracing. Next, we briefly describe the identity proof according to the above two purposes. More details are described in Section 5.

3.1.1 Proving the user's legitimacy. We assume that the user u has generated $(pub_u, priv_u)$, $(pk_{chash_u}, sk_{chash_u})$ and $CH_{id_u} = \mathcal{H}_{chash}(pk_{chash_u}, id_u, r)$, where r is the random number sampled by u.

Step 1: user registration. To prove the user's legitimacy, there must be something (similar to a certificate) that can indicate the user's legitimacy. In SkyEye, this is done through user registration. Here, we briefly introduce user registration in SkyEye. More details on user registration appear in Section 5.1.2.

As shown in Figure 3, the user u sends registration information (C_{info}, π_{info}) to the regulator, where C_{info} is the ciphertext that is the encryption of the plaintext $(pk_{chash_u}, id_u, CH_{id_u})$ under the public key pk_{reg} and π_{info} is the zk-SNARK proof that is used to prove: *"I know (sk_{chash_u}, r) which can generate pk_{chash_u} and CH_{id_u}".*

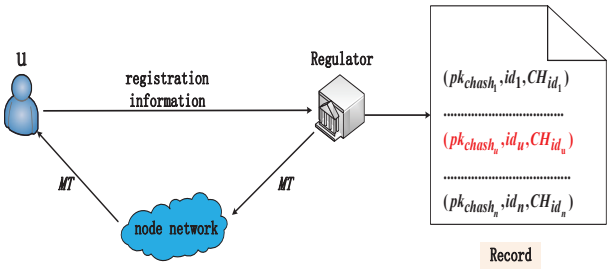

Figure 3: User registration.

If the verification of (C_{info}, π_{info}) is successful, the regulator stores $(pk_{chash_u}, id_u, CH_{id_u})$, and adds $pk_{chash_u}||CH_{id_u}$ to the Merkle tree MT. The regulator publishes the Merkle tree MT at the right time. The registration of u is successful only if $pk_{chash_u}||CH_{id_u}$ appears in the Merkle tree MT.

The Merkle tree MT can be regarded as a credential of proving the user's legitimacy. In other words, to prove the legitimacy of u, the user u must prove that $pk_{chash_u}||CH_{id_u}$ appears in the Merkle tree MT. Therefore, $proof_{id_u}$ generated by the user u must be able to prove the following.

"I know (sk_{chash_u}, id_u, r) that can generate pk_{chash_u} and CH_{id_u}, and $pk_{chash_u}||CH_{id_u}$ appears as a leaf of the Merkle tree MT with the root rt".

Step 2: establishing the binding relationship between pub_u and $proof_{id_u}$. Although $proof_{id_u}$ described above can prove the legitimacy of u, an issue remains. It can be seen from Figure 4 that $(pub_u, proof_{id_u})$ needs to be published in the node network. The adversary who has registered with the regulator can generate an identity proof $proof_{id_{adv}}$ and publish $(pub_u, proof_{id_{adv}})$ to the node work. At this time, pub_u corresponds to two different users' identity proofs (i.e., $proof_{id_u}$ and $proof_{id_{adv}}$). This presents a great obstacle to trace. We need to establish a relation between pub_u and $proof_{id_u}$ to ensure that only the user u can generate the identity proof $proof_{id_u}$ corresponding to pub_u.

The key idea behind establishing this relation is that we establish the binding relationship between $priv_u$ and $proof_{id_u}$. Because pub_u is generated by $priv_u$, there is binding relationship between pub_u and $proof_{id_u}$.

We leverage the special property of chameleon hash scheme (i.e., the user who knows the trapdoor can easily find collision) to establish the binding relationship between $priv_u$ and $proof_{id_u}$. Given the private information $priv_u$, the user u who knows sk_{chash_u} can easily find r' such that $CH_{id_u} = \mathcal{H}_{chash}(pk_{chash_u}, priv_u, r')$. To achieve the binding of $priv_u$ and $proof_{id_u}$, we require the identity proof $proof_{id_u}$ to prove the following.

- The public information pub_u is generated by the private information $priv_u$.
- I know $(sk_{chash_u}, priv_u, r')$ that generates pk_{chash_u} and CH_{id_u}.
- $pk_{chash_u}||CH_{id_u}$ appears as a leaf of a Merkle tree with the root rt.

This binding relationship between $proof_{id_u}$ and $priv_u$ ensures that only the user u who knows the private information $priv_u$ can generate the identity proof $proof_{id_u}$, and others cannot forge the identity proof corresponding to pub_u.

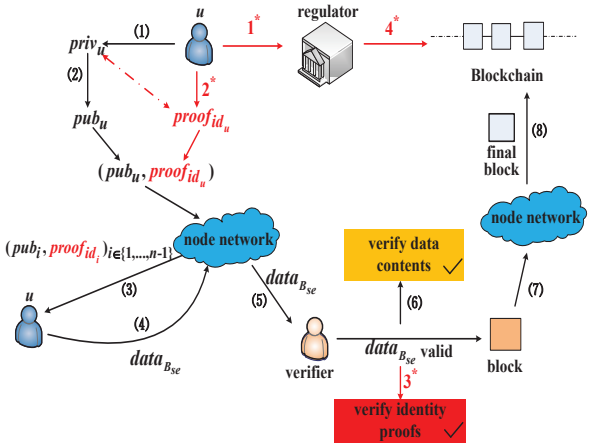

Figure 4: Overview of B_{se}. The red lines 1*, 2*, 3* and 4* represent the operations of the SkyEye scheme. The red line 1* denotes user registration. The red line 2* denotes generating identity proof. The red line 3* denotes verifying identity proof. The red line 4* denotes tracing.

Moreover, the special property of chameleon hash scheme makes the user u just register once, and then can generate identity proofs without involving the regulator.

3.1.2 Achieving tracing. To achieve tracing, we add C_{id_u}, which is the ciphertext of pk_{chash_u} under the regulator public key pk_{reg}, to $proof_{id_u}$, and require $proof_{id_u}$ to prove that the plaintext corresponding to the ciphertext C_{id_u} is pk_{chash_u}. Because the regulator has the record $(pk_{chash_i}, id_i, CH_{id_i})_{i\in\{1,\ldots,n\}}$, the regulator can obtain pk_{chash_u} through decrypting C_{id_u} and determine the identity id_u of the user u based on the record.

Remark. There are many ways to verify the user's identity in reality, such as, face recognition, identity card, or short message service (SMS) verification. Therefore, we assume there is an efficient way of verifying the user's real identity in the user registration of SkyEye.

3.2 SkyEye Overview

As can be seen from Figure 4, SkyEye's application strategy in B_s is that the user u generates the identity proof $proof_{id_u}$ corresponding to the public information pub_u so that every verifier can verify the legitimacy of u and the regulator can trace $data_{B_{se}}$.

From Figure 4, it can also be seen that the SkyEye scheme mainly has the following operations between the regulator, users, and verifiers.

- *User registration.* The user generates registration information and sends it to the regulator. The regulator is responsible for the verification of registration information.
- *Generating identity proof.* The user who registers successfully can generate the identity proof. There is a binding relationship between the identity proof and the private information.

• *Verifying identity proof.* Different from traditional verification process in the blockchain, the verifier (e.g., the miner) verifies identity proofs in addition to verifying data contents. If the data contents and identity proofs are simultaneously verified successfully, the data will be added to the block generated by the verifier.

• *Tracing.* The regulator traces the users' true identities in the blockchain data.

4 DEFINITION OF A TRACEABLE SCHEME

4.1 Definition

A traceable scheme for blockchain is a tuple of polynomial-time algorithms Π = *(Setup,* Gen_{info}, Ver_{info}, Gen_{proof}, Ver_{proof}, *Trace)* described below:

• *Setup*(λ) $\rightarrow$ *pp.* Given a security parameter λ, *Setup* returns public parameters pp. This algorithm is executed by a trusted party and is done only once. The public parameters *pp* are published and made available to all parties.

• Gen_{info}*(pp,id)* $\rightarrow$ *reginfo.* Given the public parameters pp and a user identity id, this algorithm returns the registration information *reginfo*.

• Ver_{info}*(pp,reginfo,*sk_{reg}*)* $\rightarrow$ *b.* Given the public parameters *pp*, the registration information *reginfo* and the regulator private key sk_{reg}, this algorithm returns a bit b. If verification succeeds, this algorithm returns 1; otherwise, it returns 0.

• $Gen_{proof}(pp, pub, priv, CH_{id}, pk_{chash}, sk_{chash}, r, rt, path_{id})$ $\rightarrow$ $proof_{id}$. Given the public parameters *pp*, a public/private information $(pub, priv)$, the chameleon hash value CH_{id} of *id*, the chameleon hash public/private key (pk_{chash}, sk_{chash}), the random element r for computing CH_{id}, the Merkle tree root rt, and the path $path_{id}$ from $pk_{chash}||CH_{id}$ to rt, Gen_{proof} returns the user identity proof $proof_{id}$.

• Ver_{proof}*(pp,pub,*$proof_{id}$*)* $\rightarrow$ *b.* Given the public parameters pp, the user public information *pub* and the user identity proof $proof_{id}$, Ver_{proof} returns a bit b. If the verification of $proof_{id}$ succeeds, this algorithm returns 1; otherwise, it returns 0.

• *Trace(*$\boldsymbol{data_{B_{se}}}$, sk_{reg}*)* $\rightarrow$ *ID.* Given the blockchain data $\boldsymbol{data_{B_{se}}}$ and the regulator private key sk_{reg}, *Trace* returns the identity set *ID* for $\boldsymbol{data_{B_{se}}}$.

4.2 Security

We assume that in $\boldsymbol{B_s}$ relevant cryptographic techniques (e.g., digital signatures) have been used to ensure that the blockchain data generated by the users cannot be tampered with. Therefore, the identity proofs added to the blockchain data also cannot be tampered with. We also assume that the regulator is trusted and has an efficient way of verifying user identity. Therefore, the goals of the adversary are to forge the user identity proof and to distinguish two distinct user identity proofs. The security of a traceable scheme must satisfy two properties: *identity proof indistinguishability and identity proof unforgeability.*

Definition 4.1. A traceable scheme $\prod$ = (Setup, Gen_{info}, Ver_{info}, Gen_{proof}, Ver_{proof}, Trace) is secure if it satisfies identity proof indistinguishability and identity proof unforgeability.

Below, we briefly describe each property, and defer formal definition of each property to Appendix A.

• **Identity proof indistinguishability.** This property requires that even if the adversary can adaptively induce honest parties to perform operations of his choice, the identity proof reveals no information except for some public information, such as public addresses and serial numbers. In other words, even if the adversary queries two different honest parties (one identity is id_0, and the other identity is id_1), no polynomial-time adversary can distinguish between the identity proofs $proof_{id_0}$ and $proof_{id_1}$. The meaning of this property is that if the blockchain is indistinguishable, adding the identity proofs to the blockchain data does not affect the indistinguishability of the blockchain.

• **Identity proof unforgeability.** This property requires that even if the adversary can adaptively induce honest parties to perform operations of his choice, no polynomial-time adversaries can forge the identity proof of honest parties. This property ensures that the adversary cannot forge the honest user's identity proof to create blockchain data for evading tracing.

5 CONSTRUCTION

5.1 SkyEye Construction

5.1.1 SkyEye Initialization. The public parameters *pp* created by the *Setup* algorithm include the following information: the proving/verification key (pk_{info}, vk_{info}) used to generate and verify the zk-SNARK proof π_{info} for the NP relation R_{info} (see Section 5.1.2 for details), the proving/verification key (pk_{proof}, vk_{proof}) used to generate and verify the zk-SNARK proof π_{proof} for the NP relation R_{proof} (see Section 5.1.3 for details), the regulator public key pk_{reg} for public key encryption, and the public parameters pp_{chash} of the chameleon hash scheme. Because the regulator is trusted, *Setup* algorithm is performed by the regulator. (See *Setup* algorithm in Algorithm 1 for specific operations.)

5.1.2 User Registration. As shown in Algorithm 1, the Gen_{info} algorithm is responsible for the generation of registration information and the Ver_{info} algorithm is used to verify the user's registration information.

In the Gen_{info} algorithm, a user generates the chameleon hash public-private pair (pk_{chash}, sk_{chash}) based on pp_{chash}, then computes the chameleon hash value CH_{id} of identity *id*, and stores $(id, pk_{chash}, sk_{chash}, r, CH_{id})$. At this point, the user can produce a zk-SNARK proof π_{info} for the following NP relation, which we call R_{info}:

"Given $x_{info} = (id, pk_{chash}, CH_{id})$, I know $w_{info} = (sk_{chash}, r)$ such that:

♦ The chameleon hash private key matches the chameleon hash public key: $pk_{chash} = chash\_gen(sk_{chash})$.

♦ The chameleon hash is computed correctly: $CH_{id} = \mathcal{H}_{chash}($ $pk_{chash}, id, r)$."

The Gen_{info} algorithm outputs registration information *reginfo*, which consists of the ciphertext C_{info} and zk-SNARK proof π_{info}. C_{info} is the ciphertext of x_{info} encrypted by pk_{reg}.

The verification operations in the Ver_{info} algorithm include verifying the identity *id* and verifying the zk-SNARK proof π_{info}. If the above two operations are verified successfully, the regulator stores $(pk_{chash}, id, CH_{id})$, and then publishes $pk_{chash}||CH_{id}$ stored

Algorithm 1 SkyEye Construction

Setup

Input: security parameter λ;
Output: public parameters pp;
1: construct arithmetic circuit AC_{info} for relation R_{info} at security λ;
2: construct arithmetic circuit AC_{proof} for relation R_{proof} at security λ;
3: $(pk_{info}, vk_{info}) = \mathcal{K}_{nizk}(\lambda, AC_{info})$;
4: $(pk_{proof}, vk_{proof}) = \mathcal{K}_{nizk}(\lambda, AC_{proof})$;
5: compute $pp_{enc} = \mathcal{G}_{enc}(\lambda)$;
6: compute $(pk_{reg}, sk_{reg}) = \mathcal{K}_{enc}(pp_{enc})$;
7: compute $pp_{chash} = \mathcal{G}_{chash}(\lambda)$;
8: **return** $pp = (pk_{info}, vk_{info}, pk_{proof}, vk_{proof}, pk_{reg}, pp_{chash})$;

Gen$_{info}$

Input:
public parameters pp,
user identity id;
Output:
registration information $reginfo$;
1: $(pk_{chash}, sk_{chash}) = \mathcal{K}_{chash}(pp_{chash})$;
2: randomly sample r;
3: compute $CH_{id} = \mathcal{H}_{chash}(pk_{chash}, id, r)$;
4: set $x_{info} = (id, pk_{chash}, CH_{id})$, $w_{info} = (sk_{chash}, r)$;
5: $\pi_{info} = \mathcal{P}_{nizk}(pk_{info}, x_{info}, w_{info})$;
6: set $C_{info} = \mathcal{E}_{enc}(pk_{reg}, x_{info})$;
7: store $(id, pk_{chash}, sk_{chash}, r, CH_{id})$;
8: **return** $reginfo = (C_{info}, \pi_{info})$;

Ver$_{info}$

Input:
public parameters pp,
registration information $reginfo$,
regulator private key sk_{reg};
Output: bit b;
1: parse $reginfo$ as (C_{info}, π_{info});
2: $x_{info} = \mathcal{D}_{enc}(sk_{reg}, C_{info})$;
3: parse x_{info} as $(id, pk_{chash}, CH_{id})$;
4: **if** id not valid **then**
5: **return** b=0;
6: **end if**
7: **if** $\mathcal{V}_{nizk}(vk_{info}, x_{info}, \pi_{info}) = 0$ **then**
8: **return** b=0;
9: **else**
10: store $(pk_{chash}, id, CH_{id})$;
11: publish $pk_{chash}||CH_{id}$ via the Merkle tree MT;
12: **return** b=1;
13: **end if**

Gen$_{proof}$

Input:
public parameters pp,
user public/private information $(pub, priv)$,
chameleon hash value CH_{id} of id,
chameleon hash public/private key (pk_{chash}, sk_{chash}),
random element r for computing CH_{id},
Merkle tree root rt,
path $path_{id}$ from $pk_{chash}||CH_{id}$ to rt;
Output:
user identity proof $proof_{id}$;
1: compute $r' = \mathcal{CF}_{chash}(sk_{chash}, id, priv, r)$;
2: randomly sample rn for encrypting;
3: compute $C_{id} = \mathcal{E}_{enc}(pk_{reg}, pk_{chash}, rn)$;
4: set $u_{proof} = (rt, pk_{reg}, C_{id})$;
5: set $x_{proof} = (pub, u_{proof})$,
$w_{proof} = (path_{id}, CH_{id}, sk_{chash}, pk_{chash}, priv, r', rn)$;
6: compute $\pi_{proof} = \mathcal{P}_{nizk}(pk_{proof}, x_{proof}, w_{proof})$;
7: set $proof_{id} = (u_{proof}, \pi_{proof})$;
8: **return** $proof_{id}$

Ver$_{proof}$

Input:
public parameters pp,
user public information pub,
identity proof $proof_{id}$;
Output:
bit b;
1: parse $proof_{id}$ as (u_{proof}, π_{proof})
2: set $x_{proof} = (pub, u_{proof})$
3: **if** $(\mathcal{V}_{nizk}(vk_{proof}, x_{proof}, \pi_{proof}) = 0)$ **then**
4: **return** b=0;
5: **else**
6: **return** b=1;
7: **end if**

Trace

Input:
blockchain data $\boldsymbol{data_{B_{se}}}$,
regulator private key sk_{reg};
Output:
identity set ID for $\boldsymbol{data_{B_{se}}}$;
1: set $ID = \phi$;
2: get ciphertext set $C = \{C_{id_i}\}_{i \in \{1,\ldots,n\}}$ from $\boldsymbol{data_{B_{se}}}$, where n is the number of the users' public information in $\boldsymbol{data_{B_{se}}}$;
3: **for** (each $C_{id_i} \in C$) **do**
4: compute $pk_{chash_i} = \mathcal{D}_{enc}(sk_{reg}, C_{id_i})$;
5: search $(pk_{chash}, id, CH_{id})$ records, get id_i corresponding to pk_{chash_i};
6: put id_i in ID;
7: **end for**
8: **return** ID;

in the Merkle tree MT in which the root is denoted by rt. Meanwhile, this algorithm returns 1.

5.1.3 Generating and Verifying Identity Proof. As shown in Algorithm 1, the Gen_{proof} algorithm is used to generate the identity proof for each user.

In the Gen_{proof} algorithm, assume a user has generated public/private information $(pub, priv)$. According to the known trapdoor sk_{chash}, the user can calculate a value r' such that $CH_{id} = \mathcal{H}_{chash}(pk_{chash}, priv, r')$. Next, the user computes ciphertext $C_{id} = \mathcal{E}_{enc}(pk_{reg}, pk_{chash}, rn)$, where pk_{reg} is the public key of the regulator, and rn is the random number used for encryption. Finally, the user produces a zk-SNARK proof π_{proof} for the following NP relation, which we term R_{proof}:

"Given a statement $x_{proof} = (pub, rt, pk_{reg}, C_{id})$, I know $w_{proof} = (path_{id}, CH_{id}, sk_{chash}, pk_{chash}, priv, r', rn)$ such that:

♦ The private information matches the public information: $pub = gen(priv)$.

♦ The chameleon hash private key matches the chameleon hash public key: $pk_{chash} = chash\_gen(sk_{chash})$.

♦ The chameleon hash is computed correctly: $CH_{id} = \mathcal{H}_{chash}(pk_{chash}, priv, r')$.

♦ The ciphertext C_{id} corresponds to the plaintext pk_{chash}: $C_{id} = \mathcal{E}_{enc}(pk_{reg}, pk_{chash}, rn)$.

♦ The $pk_{chash}||CH_{id}$ appears as a leaf of a Merkle tree with the root rt."

The Ver_{proof} algorithm in Algorithm 1 is used to verify the user's identity proof $proof_{id}$. The verification operation verifies the zk-SNARK proof π_{proof}. This algorithm returns 1 if and only if the above operation verifies successfully.

5.1.4 Tracing. As shown in Algorithm 1, the $Trace$ algorithm is used to trace the blockchain data $\boldsymbol{data_{B_{se}}}$. The regulator obtains pk_{chash_i} by decrypting the ciphertext C_{id_i} for each $i \in \{1, ..., n\}$, and according to the record that stores each user's chameleon hash public key, chameleon hash value, and identity, the regulator can determine the true identities of the users in the $\boldsymbol{data_{B_{se}}}$. This algorithm returns the identity set ID.

5.2 SkyEye Security

Theorem 5.1. *Assuming that the Chash scheme is collision resistant, trapdoor collision and semantic security, the NIZK scheme is perfectly zero-knowledge and simulation sound extractable, the encryption scheme Enc satisfies IND-CCA2 security, and $gen(\cdot)$ has one-wayness property. Our scheme $\prod$ = (Setup, Gen_{info}, Ver_{info}, Gen_{proof}, Ver_{proof}, Trace) described in Algorithm 1 is a secure (cf. Definition 4.1) traceable scheme.*

We provide the proof of Theorem 5.1 in Appendix B.

6 IMPLEMENTATION AND EVALUATION

6.1 Implementation

There are two main ways of generating public and private information in blockchain applications. One is through the pseudorandom function (e.g., Zerocash[7], Hawk[26]), i.e. $pub = PRF_{priv}(s)$, where PRF denotes the pseudorandom function, pub is the pseudorandom number, $priv$ is the private key used to generate pub, and s is the uniform seed. The other way is to use elliptic curve scalar multiplication (e.g., Bitcoin) to generate the public and private information, i.e., $pub = priv \cdot G$, where $priv$ is a scalar, G is a base point on the elliptical curve, and pub is a point on the elliptical curve. We use $SkyEye_H$ to represent the scheme that generates public and private information in the first way, and $SkyEye_S$ to represent the scheme that generates public and private information in the second way. We use the C++ programming language to implement the prototype of the above two different schemes based on the zk-SNARK library, libsnark[8].

There are some cryptographic building blocks in $SkyEye_H$: the pseudorandom function, chameleon hash scheme, hash function in the Merkle tree, public encryption scheme, and zk-SNARK scheme. For the chameleon hash scheme, we use the chameleon hash scheme proposed by Hugo Krawczyk and Tal Rabin[27]. For efficiency, we use the SHA256 compression function to implement the pseudorandom function and hash function in the Merkle tree, which is similar to the approach used in Zerocash[7]. We use the practical public key encryption scheme proposed by Cramer and Shoup[11], an IND-CCA2 secure public encryption scheme, as our encryption scheme. We use the scheme proposed by Parno et al.[37] as the zk-SNARK scheme. In the concrete implementation, we use the Barreto-Naehrig elliptic curve[6] that provides 128-bit security as the underlying curve of the zk-SNARK scheme. The implementation of the chameleon hash and public key encryption scheme is based on a prime field of 254 bits.

In $SkyEye_S$, the main cryptographic building block differs from the former in that the pseudorandom function is replaced by elliptic curve scalar multiplication. The chameleon hash scheme, public key encryption scheme, and zk-SNARK scheme are the same as those in the $SkyEye_H$. In the concrete implementation, we use the MNT4 elliptic curve[34] as the underlying curve of the zk-SNARK scheme. The implementation of elliptic curve scalar multiplication is based on the MNT6 elliptic curve [34]. We implement the chameleon hash scheme and public key encryption scheme in a prime field of 298 bits. To improve efficiency, in the formation of the Merkle tree, because the length of the leaf node is 298 bits, two leaf nodes together cannot form 512 bits. Therefore, the upper node is generated by the leaf node using the standard SHA256. In addition, the data length of the node above the leaf node is 256 bits, so each node that is not generated through the leaf node is generated by the SHA256 compression function.

6.2 Evaluation

We evaluate the performance of every algorithm in the two aforementioned schemes in two different configurations: configuration 1, with an Intel i5 processor and 4 GB memory laptop; and configuration 2, with an Intel i7 processor and 16 GB memory desktop machine. The depth of the Merkle tree in our evaluation is 10, 20, 30, and 34, respectively. In other words, the maximum number of users which the Merkle tree supports is 2^{10}, 2^{20}, 2^{30}, and 2^{34}. This fully meets demand, because the current global population is about 7.5 billion, and 2^{34} reaches more than 17 billion. Moreover, we evaluate the performance of the $Trace$ algorithm under the condition that there are already 1024 successfully registered users at the regulator.

Table 1: Performance of $SkyEye_H$

		Configuration 1: intel(R) core(TM) i5-2450M @2.50GHz 4GB of RAM				Configuration 2: intel(R) core(TM) i7-6700 @ 3.40GHz 16GB of RAM			
$SkyEye_H$		Tree depth							
		10	20	30	34	10	20	30	34
$Setup$	time(s)	69	114	156	175	37	61	83	94
	$\|pk_{info}\|$(KB)	480							
	$\|vk_{info}\|$(B)	574							
	$\|pk_{proof}\|$(MB)	90	149	209	231	90	149	209	231
	$\|vk_{proof}\|$(KB)	21							
Gen_{info}	time(ms)	475.7	494.3	530.1	538.5	231.5	248.0	266.2	272.4
	$\|\pi_{info}\|$(B)	287							
Ver_{info}	time(ms)	15.3	15.3	15.7	15.1	7.1	7.0	7.0	6.9
Gen_{proof}	time(s)	29	46	59	66	15	24	30	35
	$\|\pi_{proof}\|$(B)	287							
Ver_{proof}	time(ms)	10.1	10.2	10.1	10.2	4.5	4.5	4.8	4.6
$Trace$	time(ms)	0.13				0.075			

Table 2: Performance of $SkyEye_S$

		Configuration 1: intel(R) core(TM) i5-2450M @2.50GHz 4GB of RAM				Configuration 2: intel(R) core(TM) i7-6700 @ 3.40GHz 16GB of RAM			
$SkyEye_H$		Tree depth							
		10	20	30	34	10	20	30	34
$Setup$	time(s)	187	296	403	451	101	162	220	244
	$\|pk_{info}\|$(KB)	661							
	$\|vk_{info}\|$(B)	667							
	$\|pk_{proof}\|$(MB)	105	174	243	268	105	174	243	268
	$\|vk_{proof}\|$(KB)	13							
Gen_{info}	time(ms)	1398.0	1413.3	1456.3	1523.9	754.9	772.6	787.8	793.8
	$\|\pi_{info}\|$(B)	337							
Ver_{info}	time(ms)	51.8	52.4	53.4	54.4	27.3	27.8	27.0	27.3
Gen_{proof}	time(s)	56	83	109	120	30	45	58	64
	$\|\pi_{proof}\|$(B)	337							
Ver_{proof}	time(ms)	47.8	47.9	48.0	47.9	24.9	24.9	24.7	24.9
$Trace$	time(ms)	0.14				0.09			

Table 1 and Table 2 illustrate the performance results of the $Setup, Gen_{info}, Ver_{info}, Gen_{proof}, Ver_{proof}$ and $Trace$ algorithms in $SkyEye_H$ and $SkyEye_S$, respectively (the time in the two tables is the average of 10 runs per algorithm). In the two tables, time represents the running time of the algorithm, and $|\cdot|$ represents the data length. For example, the $|pk_{info}|$ represents the length of the proving key in the registration. Without loss of generality, using an i7 processor, a 16 GB memory desktop machine, and with a tree depth of 34 in Table 1, we can obtain the results of the $SkyEye_H$ scheme:

• $Setup$ algorithm takes 94 s. The size of the proving key and verification key used for user registration are 480 KB and 574 B, respectively. And the size of the proving key and verification key used for user identity proof are 231 MB and 21 KB, respectively.

• Gen_{info} requires 272.4 ms, and the size of the zk-SNARK proof π_{info} is 287 B.

• Ver_{info} algorithm takes 6.9 ms.

• Gen_{proof} algorithm takes 35 s to generate a user's identity proof, and the size of the zk-SNARK proof π_{proof} is 287 B.

• Ver_{proof} algorithm takes 4.6 ms.

• $Trace$ algorithm takes 0.075 ms to trace a user's identity.

The tables reveal the following:

• In each configuration, the time required for verification by the regulator and the verifiers is small and does not substantially change as the depth of the tree changes. As shown in Table 1, the

regulator takes approximately 15 ms to verify the user registration information in configuration 1 and approximately 7 ms in configuration 2; and the time taken by a verifier to verify the user identity proof is approximately 10 ms in configuration 1 and approximately 5 ms in configuration 2. From Table 2, we can observe that the time taken for verifying the user registration information is approximately 53 ms in configuration 1 and approximately 28 ms in configuration 2; and the time taken for verifying the user identity proof is approximately 48 ms in configuration 1 and approximately 25 ms in configuration 2.

• Not all of the information in SkyEye must be on-chain. Only the information $proof_{id}$ generated by the Gen_{proof} algorithm is added to the user data. Furthermore, the size of the user's proof π_{proof} in the $proof_{id}$ is dominant. As can be observed from the two tables, the length of the zk-SNARK proof π_{proof} will not change as the configuration environment and tree depth change. The size of π_{proof} is small, and the length is 287 B in $SkyEye_H$ and 337 B in $SkyEye_S$.

7 POTENTIAL APPLICATIONS

The SkyEye scheme provides an alternative traceable strategy for the blockchain applications that belong to $\boldsymbol{B_s}$. If a blockchain application that does not belong to $\boldsymbol{B_s}$ wants to use SkyEye, this application can modify some rules to make it belong to $\boldsymbol{B_s}$.

In this section, we briefly describe how to apply SkyEye to some applications: Bitcoin[35], Ethereum[41], and RSCoin[13].

7.1 Bitcoin

In decentralized cryptocurrencies, Bitcoin is undoubtedly the most eye-catching one. It has achieved widespread adoption, and many alt-coins[1, 25] are derived from Bitcoin. Below, we briefly describe the application of SkyEye in Bitcoin.

1) Certifiable user. It is well known that users in Bitcoin are anonymous (pseudonyms). However, users wish to trade with certifiable merchants in many cases, which is more secure and more assured. At this point, SkyEye can be used in Bitcoin to allow the merchants that need to be certified to register with the regulator, and when merchants open their addresses, they also disclose their identity proofs for others to use. In this way, the miner sometimes needs to verify the identity proof. Finally, the public address with an identity proof in the blockchain indicates a certifiable user.

2) Full tracing. Because of the anonymity in Bitcoin, it is difficult to trace some illegal activities (such as money laundering and ransomware), which makes many countries ban Bitcoin transactions. If one day, Bitcoin demands tracing user identity, our SkyEye scheme can be applied. Every user who wishes to use Bitcoin must register with the regulator. Users must add identity proofs when generating a transaction, and because Bitcoin transactions are linkable, the user only needs to add identity proofs to the outputs of the transaction. When the miner verifies the transaction, he must also verify the identity proofs. Ultimately, the regulator can fully trace all data in the Bitcoin blockchain.

Both of the above strategies need to modify the underlying script code and the underlying rules in Bitcoin to support the application of the SkyEye scheme.

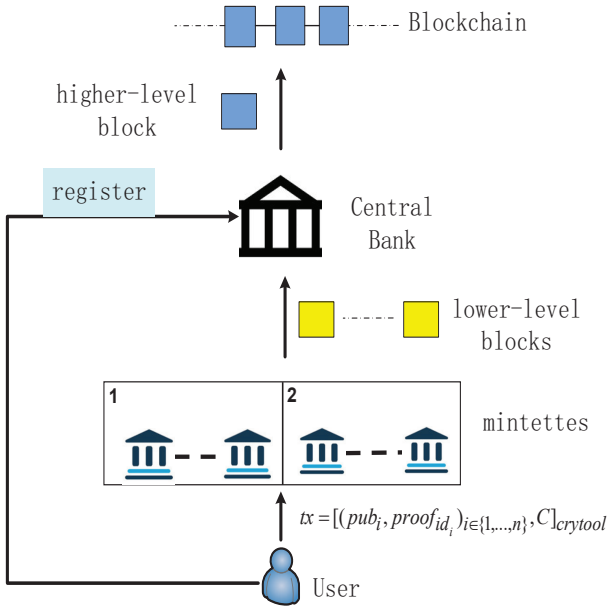

Figure 5: RSCoin with SkyEye

7.2 Ethereum

Smart contract was first proposed by Nick Szabo[39]. However, because of the absence of a credible execution environment and other technical limitations at the time, smart contract was not applied to the real world. Blockchain technology provides a natural, trusted execution environment for smart contract, and makes the application of smart contract possible. Ethereum is one of the representative smart contract platforms. Next, we describe how to use SkyEye in Ethereum.

1) Providing a tracing strategy for some applications. When an application in Ethereum demands tracing user identity, SkyEye can provide tracing functionality for this application. The Ethereum transaction has a data area, which allows users to add data. Therefore, the identity proofs can be appended to the data area. When the transaction triggers application contract code execution, the application contract code will first call the SkyEye contract code to verify identity proofs in this transaction. When this transaction is in the blockchain, the regulator can determine the identities in this transaction.

2) Tracing the entire Ethereum platform. This must modify the underlying rules of Ethereum that make only users registered with the regulator can use the Ethereum platform to develop contract codes or conduct transactions. The regulator can trace the data in the Ethereum platform using the $Trace$ algorithm.

7.3 RSCoin

Although existing private digital currency systems (such as Bitcoin) exhibit advantages such as cost savings, transaction transparency and high security, their problems (e.g., low transaction throughput, resource consumption, and difficulty in supervision) have severely restricted the application of digital currency. To make better use of the advantages of digital currency and to prevent the

risks and harms caused by private digital currency, many countries in the world such as China, Britain, and America have performed studies on digital currency, striving to issue legal digital currency as soon as possible to enhance their international competitiveness. The representative is RSCoin[13], the first digital currency framework.

The SkyEye can combine with the RSCoin framework to enable the central bank to trace transactions in the blockchain. As can be seen from Figure 5, the user first calls the Gen_{info} algorithm to register with the central bank. After the central bank passes the verification, the user can generate the transaction tx, which contains the data body and the identity proofs (each user generates an identity proof using the Gen_{proof} algorithm) and send tx to the mintettes that have been certified by the central bank. Mintettes are divided into two groups: group 1 verifies whether the user's input address is legal (such as whether the input belongs to the UTXO collection). In addition to the original verification of RScoin, group 1 must also verify the identity proofs in the transaction tx. Finally, the group 2 provides several low-level blocks to the central bank. The central bank merges the blocks to produce a high-level block and adds it to the blockchain. At this point, the central bank can obtain the identities of the users using the *Trace* algorithm in SkyEye to complete the tracing of the blockchain.

8 DISCUSSION AND FUTURE WORK

In SkyEye, the centralization of the regulator is a major issue. The regulator can arbitrarily trace the identity of blockchain data without any restrictions and oversight.

From the data tracing process of the regulator, it can be seen that the regulator must first use its private key sk_{reg} to decrypt the ciphertext of each user's chameleon hash public key in the blockchain data. Therefore, we can restrict the regulator through the distributed key generation (DKG) protocol[22]. Specifically, the public/private key pair (pk_{reg}, sk_{reg}) is generated by a committee with a threshold of t through the DKG protocol. In this way, pk_{reg} is made public, and each committee member has a share of sk_{reg}. The regulator submits the data and tracing evidence to the committee. If at least t+1 members of the committee accept the data and tracing evidence, the regulator will obtain sk_{reg} from the committee.

However, this approach does not completely restrict the regulator. Even if the committee regularly updates the public/private key pair, as long as the regulator obtains the private key sk_{reg} in a cycle, it can trace not only the data submitted to the committee, but also all user data in this cycle. In future work, we will consider how to restrict the regulator to make the regulator only trace the data submitted to the committee.

9 RELATED WORK

Blockchain research focuses primarily on enhancing blockchain privacy protection [7, 12, 33, 38], improving blockchain scalability [16, 31, 42], analyzing blockchain security[17–19, 30], and applying blockchain to other areas[28, 29, 32, 40]. However, research on traceable mechanisms is limited.

Narula, Vasquez, and Virza proposed zkLedger[36], the first distributed ledger system, that provides strong privacy protection, public verifiability, and practical auditing. zkLedger uses table construction in the ledger. Each user identity corresponds to each column in the ledger. Therefore, the regulator can determine every user identity through the ledger. However, this traceable mechanism in zkLedger cannot be applied to environments with a large number of users and is used only for auditing digital asset transactions over some banks.

Defrawy and Lampkins[15] proposed a proactively-private digital currency (PDC) scheme that can provide privacy-preserving and accountability. In their scheme, the ledger is kept by a group of ledger servers. Every ledger sever has a balance ledger that contains a share of every user identity. Therefore, the regulator can determine every user identity through those ledger servers. However, their traceable mechanism does not seem to have been extended to other applications.

Ateniese and Faonio[4] constructed a scheme that provides certified Bitcoin addresses to enable Bitcoin users to trade with certifiable users authenticated by the trusted certificate authority. The regulator can determine every user identity through the authority. However, if a user wants to use a new certified address for each transaction, the user must contact the certificate authority to obtain a certified address. This reduces the efficiency of the entire system and exerts considerable pressure on the certificate authority when the number of users is large. Moreover, their approach only applies to Bitcoin.

Garman et al.[20] designed new decentralized anonymous payment (DAP) systems to address the regulatory issue by adding privacy preseving policy-enforcement mechanisms that guarantee regulatory compliance, allow selective user tracing, and admit tracing of tainted coins. The regulator can determine every user identity through the identity escrow policy. However, the DAP system are based on Zerocash[7].

The traceable mechanisms proposed above can only be applied to specific application environments and do not seem to have been extended to other applications. We propose SkyEye, a traceable scheme for blockchain. Our scheme can be applied to a class of blockchain applications, which is denoted by $\boldsymbol{B_S}$.

10 CONCLUSION

In this paper, we design SkyEye, a traceable scheme for blockchain. SkyEye can be applied to the blockchain applications that satisfy the following conditions: (I) The users have public and private information, where the public information is generated by the private information; (II) The users' public information is disclosed in the blockchain data. SkyEye just requires the user to register only once, and enables the regulator to trace users' identities. Moreover, we implement two different SkyEye prototypes: $SkyEye_H$ and $SkyEye_S$. Our evaluation results show that even if the number of users is very large, the registration information and identity proof are verified quickly.

ACKNOWLEDGMENTS

This work was supported in part by the National Key R&D Program of China (2017YFB0802500), Beijing Municipal Science and Technology Project (No. Z191100007119007), and Shandong province major science and technology innovation project (2019JZZY020129).

REFERENCES

[1] [n.d.]. https://litecoin.org/.

[2] [n.d.]. https://en.wikipedia.org/wiki/Ransomware.

[3] Giuseppe Ateniese and Breno de Medeiros. 2004. On the Key Exposure Problem in Chameleon Hashes. In *Security in Communication Networks, 4th International Conference, SCN 2004, Amalfi, Italy, September 8-10, 2004, Revised Selected Papers*. 165–179. https://doi.org/10.1007/978-3-540-30598-9_12

[4] Giuseppe Ateniese, Antonio Faonio, Bernardo Magri, and Breno de Medeiros. 2014. Certified Bitcoins. In *Applied Cryptography and Network Security - 12th International Conference, ACNS 2014, Lausanne, Switzerland, June 10-13, 2014. Proceedings*. 80–96. https://doi.org/10.1007/978-3-319-07536-5_6

[5] Christian Badertscher, Peter Gazi, Aggelos Kiayias, Alexander Russell, and Vassilis Zikas. 2018. Ouroboros Genesis: Composable Proof-of-Stake Blockchains with Dynamic Availability. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, CCS 2018, Toronto, ON, Canada, October 15-19, 2018*. 913–930. https://doi.org/10.1145/3243734.3243848

[6] Paulo S. L. M. Barreto and Michael Naehrig. 2005. Pairing-Friendly Elliptic Curves of Prime Order. In *Selected Areas in Cryptography, 12th International Workshop, SAC 2005, Kingston, ON, Canada, August 11-12, 2005, Revised Selected Papers*. 319–331. https://doi.org/10.1007/11693383_22

[7] Eli Ben-Sasson, Alessandro Chiesa, Christina Garman, Matthew Green, Ian Miers, Eran Tromer, and Madars Virza. 2014. Zerocash: Decentralized Anonymous Payments from Bitcoin. In *2014 IEEE Symposium on Security and Privacy, SP 2014, Berkeley, CA, USA, May 18-21, 2014*. 459–474. https://doi.org/10.1109/SP.2014.36

[8] Eli Ben-Sasson, Alessandro Chiesa, Eran Tromer, and Madars Virza. 2014. Succinct Non-Interactive Zero Knowledge for a von Neumann Architecture. In *Proceedings of the 23rd USENIX Security Symposium, San Diego, CA, USA, August 20-22, 2014*. 781–796. https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/ben-sasson

[9] Nir Bitansky, Alessandro Chiesa, Yuval Ishai, Rafail Ostrovsky, and Omer Paneth. 2013. Succinct Non-interactive Arguments via Linear Interactive Proofs. In *Theory of Cryptography - 10th Theory of Cryptography Conference, TCC 2013, Tokyo, Japan, March 3-6, 2013. Proceedings*. 315–333. https://doi.org/10.1007/978-3-642-36594-2_18

[10] Miguel Castro and Barbara Liskov. 1999. Practical Byzantine Fault Tolerance. In *Proceedings of the Third USENIX Symposium on Operating Systems Design and Implementation (OSDI), New Orleans, Louisiana, USA, February 22-25, 1999*. 173–186. https://doi.org/10.1145/296806.296824

[11] Ronald Cramer and Victor Shoup. 1998. A Practical Public Key Cryptosystem Provably Secure Against Adaptive Chosen Ciphertext Attack. In *Advances in Cryptology - CRYPTO '98, 18th Annual International Cryptology Conference, Santa Barbara, California, USA, August 23-27, 1998, Proceedings*. 13–25. https://doi.org/10.1007/BFb0055717

[12] George Danezis, Cédric Fournet, Markulf Kohlweiss, and Bryan Parno. 2013. Pinocchio coin: building zerocoin from a succinct pairing-based proof system. In *PETShop'13, Proceedings of the 2013 ACM Workshop on Language Support for Privacy-Enhancing Technologies, Co-located with CCS 2013, November 4, 2013, Berlin, Germany*. 27–30. https://doi.org/10.1145/2517872.2517878

[13] George Danezis and Sarah Meiklejohn. 2016. Centrally Banked Cryptocurrencies. In *23rd Annual Network and Distributed System Security Symposium, NDSS 2016, San Diego, California, USA, February 21-24, 2016*. http://wp.internetsociety.org/ndss/wp-content/uploads/sites/25/2017/09/centrally-banked-cryptocurrencies.pdf

[14] Bernardo David, Peter Gazi, Aggelos Kiayias, and Alexander Russell. 2018. Ouroboros Praos: An Adaptively-Secure, Semi-synchronous Proof-of-Stake Blockchain. In *Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings, Part II*. 66–98. https://doi.org/10.1007/978-3-319-78375-8_3

[15] Karim El Defrawy and Joshua Lampkins. 2014. Founding Digital Currency on Secure Computation. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, November 3-7, 2014*. 1–14. https://doi.org/10.1145/2660267.2660293

[16] Stefan Dziembowski, Sebastian Faust, and Kristina Hostáková. 2018. General State Channel Networks. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, CCS 2018, Toronto, ON, Canada, October 15-19, 2018*. 949–966. https://doi.org/10.1145/3243734.3243856

[17] Ittay Eyal. 2015. The Miner's Dilemma. In *2015 IEEE Symposium on Security and Privacy, SP 2015, San Jose, CA, USA, May 17-21, 2015*. 89–103. https://doi.org/10.1109/SP.2015.13

[18] Ittay Eyal and Emin Gün Sirer. 2014. Majority Is Not Enough: Bitcoin Mining Is Vulnerable. In *Financial Cryptography and Data Security - 18th International Conference, FC 2014, Christ Church, Barbados, March 3-7, 2014, Revised Selected Papers*. 436–454. https://doi.org/10.1007/978-3-662-45472-5_28

[19] Juan A. Garay, Aggelos Kiayias, and Nikos Leonardos. 2015. The Bitcoin Backbone Protocol: Analysis and Applications. In *Advances in Cryptology - EUROCRYPT 2015 - 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Sofia, Bulgaria, April 26-30, 2015, Proceedings, Part II*. 281–310. https://doi.org/10.1007/978-3-662-46803-6_10

[20] Christina Garman, Matthew Green, and Ian Miers. 2016. Accountable Privacy for Decentralized Anonymous Payments. In *Financial Cryptography and Data Security - 20th International Conference, FC 2016, Christ Church, Barbados, February 22-26, 2016, Revised Selected Papers*. 81–98. https://doi.org/10.1007/978-3-662-54970-4_5

[21] Rosario Gennaro, Craig Gentry, Bryan Parno, and Mariana Raykova. 2013. Quadratic Span Programs and Succinct NIZKs without PCPs. In *Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings*. 626–645. https://doi.org/10.1007/978-3-642-38348-9_37

[22] Rosario Gennaro, Stanislaw Jarecki, Hugo Krawczyk, and Tal Rabin. 1999. Secure Distributed Key Generation for Discrete-Log Based Cryptosystems. In *Advances in Cryptology - EUROCRYPT '99, International Conference on the Theory and Application of Cryptographic Techniques, Prague, Czech Republic, May 2-6, 1999, Proceeding*. 295–310. https://doi.org/10.1007/3-540-48910-X_21

[23] Thomas Hardjono, Ned Smith, and Alex Sandy Pentland. 2014. Anonymous identities for permissioned blockchains. https://petertodd.org/assets/2016-04-21/MIT-ChainAnchor-DRAFT.pdf.

[24] Aggelos Kiayias, Alexander Russell, Bernardo David, and Roman Oliynykov. 2017. Ouroboros: A Provably Secure Proof-of-Stake Blockchain Protocol. In *Advances in Cryptology - CRYPTO 2017 - 37th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 20-24, 2017, Proceedings, Part I*. 357–388. https://doi.org/10.1007/978-3-319-63688-7_12

[25] Sunny King and Scott Nadal. 2012. PPCoin: Peer-to-Peer Crypto-Currency with Proof-of-Stake. https://peercoin.net/assets/paper/peercoin-paper.pdf.

[26] Ahmed E. Kosba, Andrew Miller, Elaine Shi, Zikai Wen, and Charalampos Papamanthou. 2016. Hawk: The Blockchain Model of Cryptography and Privacy-Preserving Smart Contracts. In *IEEE Symposium on Security and Privacy, SP 2016, San Jose, CA, USA, May 22-26, 2016*. 839–858. https://doi.org/10.1109/SP.2016.55

[27] Hugo Krawczyk and Tal Rabin. 1998. Chameleon Hashing and Signatures. *IACR Cryptology ePrint Archive* 1998 (1998), 10. http://eprint.iacr.org/1998/010

[28] Ranjit Kumaresan and Iddo Bentov. 2016. Amortizing Secure Computation with Penalties. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016*. 418–429. https://doi.org/10.1145/2976749.2978424

[29] Ranjit Kumaresan, Vinod Vaikuntanathan, and Prashant Nalini Vasudevan. 2016. Improvements to Secure Computation with Penalties. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016*. 406–417. https://doi.org/10.1145/2976749.2978421

[30] Yujin Kwon, Dohyun Kim, Yunmok Son, Eugene Y. Vasserman, and Yongdae Kim. 2017. Be Selfish and Avoid Dilemmas: Fork After Withholding (FAW) Attacks on Bitcoin. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS 2017, Dallas, TX, USA, October 30 - November 03, 2017*. 195–209. https://doi.org/10.1145/3133956.3134019

[31] Loi Luu, Viswesh Narayanan, Chaodong Zheng, Kunal Baweja, Seth Gilbert, and Prateek Saxena. 2016. A Secure Sharding Protocol For Open Blockchains. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016*. 17–30. https://doi.org/10.1145/2976749.2978389

[32] Stephanos Matsumoto and Raphael M. Reischuk. 2017. IKP: Turning a PKI Around with Decentralized Automated Incentives. In *2017 IEEE Symposium on Security and Privacy, SP 2017, San Jose, CA, USA, May 22-26, 2017*. 410–426. https://doi.org/10.1109/SP.2017.57

[33] Ian Miers, Christina Garman, Matthew Green, and Aviel D. Rubin. 2013. Zerocoin: Anonymous Distributed E-Cash from Bitcoin. In *2013 IEEE Symposium on Security and Privacy, SP 2013, Berkeley, CA, USA, May 19-22, 2013*. 397–411. https://doi.org/10.1109/SP.2013.34

[34] A. Miyaji M. Nakabayashi and S. Takano. 2001. New explicit conditions of elliptic curve traces for FR-reduction. *IEICE Transactions on Fundamentals* 84, 5 (2001), 1234–1243.

[35] Satoshi Nakamoto. 2009. Bitcoin: A Peer-to-Peer Electronic Cash System. https://bitcoin.org/bitcoin.pdf.

[36] Neha Narula, Willy Vasquez, and Madars Virza. 2018. zkLedger: Privacy-Preserving Auditing for Distributed Ledgers. In *15th USENIX Symposium on Networked Systems Design and Implementation, NSDI 2018, Renton, WA, USA, April 9-11, 2018*. 65–80. https://www.usenix.org/conference/nsdi18/presentation/narula

[37] Bryan Parno, Jon Howell, Craig Gentry, and Mariana Raykova. 2013. Pinocchio: Nearly Practical Verifiable Computation. In *2013 IEEE Symposium on Security and Privacy, SP 2013, Berkeley, CA, USA, May 19-22, 2013*. 238–252. https://doi.org/10.1109/SP.2013.47

[38] Tim Ruffing, Pedro Moreno-Sanchez, and Aniket Kate. 2017. P2P Mixing and Unlinkable Bitcoin Transactions. In *24th Annual Network and Distributed System Security Symposium, NDSS 2017, San Diego, California, USA, February 26 - March 1, 2017*. https://www.ndss-symposium.org/ndss2017/ndss-2017-programme/p2p-mixing-and-unlinkable-bitcoin-transactions/

[39] Nick Szabo. 1997. The idea of smart contracts. http://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/idea.html.

[40] Alin Tomescu and Srinivas Devadas. 2017. Catena: Efficient Non-equivocation via Bitcoin. In *2017 IEEE Symposium on Security and Privacy, SP 2017, San Jose, CA, USA, May 22-26, 2017*. 393–409. https://doi.org/10.1109/SP.2017.19

[41] GAVIN WOOD. Accessed: 2016-05-15. Ethereum: A secure decentralized transaction ledger. http://gavwood.com/paper.pdf.

[42] Mahdi Zamani, Mahnush Movahedi, and Mariana Raykova. 2018. RapidChain: Scaling Blockchain via Full Sharding. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, CCS 2018, Toronto, ON, Canada, October 15-19, 2018*. 931–948. https://doi.org/10.1145/3243734.3243853

A SECURITY OF THE TRACEABLE SCHEME

We describe *identity proof indistinguishability* and *identity proof unforgeability*. Every property is formalized as an experiment between an adversary $\mathcal{A}$ and a challenger C. The behavior of the honest user with identity id is realized by the oracle O_{id}, and the behavior of the regulator is realized by the oracle O_{reg}. We assume that the honest users and adversary in the experiment have already registered successfully in the regulator, i.e., they can generate any identity proof. Below, we describe how O_{id} and O_{reg} work.

Oracles O_{id} and O_{reg} are initialized by challenger C using the public parameters pp. O_{reg} stores: (1) **Record**, a set of information used to trace true identities of all registered users; (2) the encryption public/private key pair (pk_{reg}, sk_{reg}). O_{reg} accepts different queries, which are described below:

• $Q = (judge, proof_{id_1}, proof_{id_2})$.

O_{reg} determines whether $proof_{id_1}$ and $proof_{id_2}$ belong to the same user, and sends the result to the inquirer.

• $Q = (chashset, proof_{id})$.

O_{reg} sends the chameleon hash set P_{chash} to the inquirer, where P_{chash} includes the chameleon hash value of the user who generates $proof_{id}$.

O_{id} stores: (1) **RegPriInfo**, the secret information used to generate registration information; (2) **IdProof**, a set of identity proofs generated by the user whose identity is id; (3)**IdProofPriInfo**, the set of evidence that the user uses to generate the identity proofs. The oracle O_{id} accepts different queries, which are described below:

• $Q = (genidproof)$. The adversary is not aware of the private information $priv$. The oracle O_{id} first randomly selects $priv$, and then generates the public information pub. Finally, the oracle O_{id} calls the Gen_{proof} algorithm to generate the identity proof $proof_{id}$, and sends $(pub, proof_{id})$ to the inquirer.

• $Q = (genidproof, priv)$.

The adversary knows the private information $priv$, and the oracle O_{id} uses the $priv$ selected by the adversary to generate the public information pub and then calls the Gen_{proof} algorithm to generate the identity proof $proof_{id}$. Finally, O_{id} sends $(pub, proof_{id})$ to the inquirer.

• $Q = (genidproof, pub_i)$.

Here, $pub_i \in T_{pub}$, and $T_{pub} = \{pub_i\}_{i\in\{1,...,n\}}$ is the public information set of the user whose identity is id. The oracle O_{id} calls the Gen_{proof} algorithm to generate the identity proof $proof_{id}$, and sends $(pub_i, proof_{id})$ to the inquirer.

A.1 Identity proof indistinguishability

This property is formalized by $Exp^{IDP-IND}_{\mathcal{A},\Pi}(\lambda)$, which is shown below:

1. The challenger C randomly samples $b \in \{0,1\}$, gets pp by running Setup(λ), and sends pp to adversary $\mathcal{A}$. Next, C initializes two separate oracles O_{id_0} and O_{id_1}.

2. At each query phase, the adversary $\mathcal{A}$ issues a pair of queries (Q, Q'), where (Q, Q') is one of the following::

• Q and Q' are both genidproof queries. C forwards Q to O_{id_0}, and forwards Q' to O_{id_1}. C replies to $\mathcal{A}$ with $((pub_b, proof_{id_b}), (pub_{1-b}, proof_{id_{1-b}}))$, which is the two oracle answer.

• $\{Q, Q'\} = \{(genidproof, priv), (genidproof, priv')\}$, where $priv = priv'$. C forwards Q to O_{id_0}, and forwards Q' to O_{id_1}. C replies to $\mathcal{A}$ with $((pub, proof_{id_b}), (pub, proof_{id_{1-b}}))$, which is the two oracle answer.

3. At the end of the query, $\mathcal{A}$ sends C a guess $b' \in \{0,1\}$. If $b = b'$, C outputs 1; otherwise, C outputs 0.

Identity proof indistinguishability requires that the adversary $\mathcal{A}$ wins the above experiment with only negligible probability. Next, we formally define this property.

Definition A.1. A traceable scheme $\prod$ satisfies identity proof indistinguishability if for all probabilistic polynomial-time adversaries $\mathcal{A}$, there is a negligible function $negl(\cdot)$ such that

$$Adv^{IDP-IND}_{A,\Pi} \leq negl(\lambda), \tag{1}$$

where $Adv^{IDP-IND}_{A,\Pi} = Pr[Exp^{IDP-IND}_{\mathcal{A},\Pi}(\lambda) = 1] - 1/2$ is $\mathcal{A}$'s advantage in the experiment $Exp^{IDP-IND}_{\mathcal{A},\Pi}(\lambda)$.

A.2 Identity proof unforgeability

This property is formalized by $Exp^{IDP-UNF}_{\mathcal{A},\Pi}(\lambda)$, which is shown below:

1. The challenger C obtains pp by running Setup(λ), and sends pp to adversary $\mathcal{A}$. Next, C initializes two separate oracles O_{id} and O_{reg}. Let $T_{pub} = \{pub_1, ..., pub_n\}$ be the public information set for the user whose identity is id.

2. The adversary $\mathcal{A}$ issues queries $q_1, ..., q_m$, where q_i is $(genidproof, pub_i)$, and $pub_i \in T_{pub}$. C forwards Q to O_{id}, C replies to $\mathcal{A}$ with $(pub_i, proof_{id})$, which is the oracle O_{id}'s answer.

3. At the end of the query, let $P = \{proof_1, ..., proof_m\}$ is the identity proof set that is generated by O_{id}. $\mathcal{A}$ sends $(pub^*, proof^*_{id})$ to C. C checks as follows:

• If $proof^*_{id} \notin P \wedge Ver_{proof}(pp, pub^*, proof^*_{id}) = 1$, C proceeds as follows; otherwise it aborts.

• C sends $(judge, proof_{id^*}, proof_i)$ to O_{reg}, where $i \in [1, m]$. If $proof_{id^*}$ and $proof_i$ belong to the user whose the identity is id, O_{reg} sends $c = 1$ to C; otherwise it returns $c = 0$.

If $proof^*_{id} \notin P \wedge Ver_{proof}(pp, pub^*, proof^*_{id}) = 1 \wedge c = 1$, C outputs 1; otherwise, C outputs 0.

The adversary $\mathcal{A}$ wins the above experiment if $proof^*_{id}$ such that: (i) $proof^*_{id} \notin P$; (ii) $Ver_{proof}(pp, pub^*, proof^*_{id}) = 1$; (iii) $proof_{id^*}$ belongs to the user whose identity is id. In other words, $\mathcal{A}$ can forge the identity proof of honest parties. Identity proof unforgeability requires that the adversary wins the above experiment with only negligible probability. Next, we formally define this property.

Definition A.2. A traceable scheme $\prod$ satisfies identity proof unforgeability if for all probabilistic polynomial time adversaries $\mathcal{A}$, there is a negligible function $negl(\cdot)$ such that

$$Adv_{A,\Pi}^{IDP-UNF} \leq negl(\lambda), \tag{2}$$

where $Adv_{A,\Pi}^{IDP-UNF} = Pr[Exp_{\mathcal{A},\Pi}^{IDP-UNF}(\lambda) = 1] - 1/2$ is $\mathcal{A}$'s advantage in the experiment $Exp_{\mathcal{A},\Pi}^{IDP-UNF}(\lambda)$.

B PROOF OF THEOREM 5.1

B.1 Proof of identity proof indistinguishability

Theorem B.1. *Assuming that the NIZK scheme is perfectly zero-knowledge and simulation sound extractable, the encryption scheme Enc satisfies IND-CCA2 security, then, our scheme $\prod$ described in Algorithm 1 satisfies identity proof indistinguishability.*

We prove the Theorem B.1 through a sequence of hybrid experiments. Let q_m be the number of queries issued by the adversary $\mathcal{A}$.

$\boldsymbol{exp_{real}}$. The experiment exp_{real} is the same as the experiment $Exp_{\mathcal{A},\Pi}^{IDP-IND}(\lambda)$.

$\boldsymbol{exp}_1$. This experiment is the same as the experiment exp_{real} except that the challenger C simulates the $NIZK$. More precisely, C calls a polynomial-time simulator $S_{nizk}(\lambda, AC_{proof})$ to obtain $(pk_{proof}, vk_{proof}, tra)$, where tra is the trapdoor, instead of invoking $\mathcal{K}_{nizk}(\lambda, AC_{proof})$. When an oracle O_{id} sends a NIZK proof π_{proof} to C, C replaces the real proof with a simulated proof by invoking $S_{nizk}(pk_{proof}, x_{proof}, tra)$, without using the witness. Because the NIZK scheme is perfectly zero-knowledge, the distribution of the simulated π_{proof} is identical to that of the proof computed in exp_{real}. Therefore, $Adv_{exp_{real}} = Adv_{exp_1}$.

$\boldsymbol{exp_{final}}$. The experiment exp_{final} is the same as the experiment exp_1 except that the challenger C replaces C_{id} in $proof_{id}$ by encrypting a random string. More precisely, when an oracle O_{id} sends an identity proof $proof_{id}$ to C, C replaces the C_{id} with a C'_{id} generated by $\mathcal{E}_{enc}(pk_{reg}, r, rn)$, where r is a random string sampled uniformly from the plaintext space of the encryption scheme. Because the responses to the adversary $\mathcal{A}$ in exp_{final} are independent of the bit b. Therefore, $Adv_{exp_{final}} = 0$ in the experiment exp_{final}.

Next, we prove that no polynomial-time adversary can distinguish exp_1 from exp_{final} except with negligible probability(see below lemma).

Lemma 1. *After q_m queries, $|Adv_{exp_{final}} - Adv_{exp_1}| \leq q_m \cdot Adv_{enc}$, where Adv_{enc} denotes the adversary's advantage in the IND-CCA2 experiment.*

Proof sketch. We construct an algorithm $\mathcal{B}$, using $\mathcal{A}$ as a subroutine, to win the IND-CCA2 experiment.

Let $\epsilon = Adv_{exp_{final}} - Adv_{exp_1}$. For some $i \in \{1, ..., q_m\}$, when $\mathcal{A}$ issues an i-th query, $\mathcal{B}$ use the same method as $\boldsymbol{exp}_1$ to generate $proof_{id}$ except for the ciphertext C_{id} generation method. $\mathcal{B}$ chooses a random string r that has the same length as plaintext m corresponding to C_{id}. $\mathcal{B}$ sends $(m_0, m_1) = (m, r)$ to the IND-CCA2 challenger and receives $C^* = \mathcal{E}_{enc}(pk_{reg}, m_{\overline{b}}, rn)$, where $\overline{b}$ is the bit chosen by the IND-CCA2 challenger. $\mathcal{B}$ replaces C_{id} included in $proof_{id}$ with C^*. $\mathcal{B}$ return b', which $\mathcal{A}$ outputs as the guess in the IND-CCA2 experiment. We know that when $\overline{b} = 0$, $\mathcal{A}$'s view of the interaction is distributed identically to that of exp_1. And when $\overline{b} = 1$, $\mathcal{A}$'s view represents the exp_1 in which one ciphertext C_{id} has been replaced. Based on a standard hybrid argument over each of the q_m ciphertexts, we can conclude that over the randomness of the experiment, $\mathcal{B}$ must succeed in the IND-CCA2 experiment with the advantage of at least ϵ/q_m. Therefore, $|Adv_{exp_{final}} - Adv_{exp_1}| \leq q_m \cdot Adv_{enc}$.

B.2 Proof of identity proof unforgeability

Theorem B.2. *Assuming that the Chash scheme is collision resistant, trapdoor collision and semantic security, the NIZK scheme is perfectly zero-knowledge and simulation sound extractable, $gen(\cdot)$ is one-wayness, then, our scheme $\prod$ described in Algorithm 1 satisfies identity proof unforgeability.*

From experiment $Exp_{\mathcal{A},\Pi}^{IDP-UNF}(\lambda)$, we can observe that $\mathcal{A}$ succeeds only if it outputs $(pub^*, proof_{id}^*)$ such that: (i) $proof_{id}^* \notin P$; (ii) $Ver_{proof}(pp, pub^*, proof_{id}^*) = 1$; (iii) $proof_{id^*}$ belongs to the user whose identity is id. We define the two disjoint events which $\mathcal{A}$ succeeds: (i) $Event$, $\mathcal{A}$ succeeds, and $pub^* \in T_{pub}$; (ii) $\overline{Event}$, $\mathcal{A}$ succeeds, and $pub^* \notin T_{pub}$.

Obviously, $Adv_{A,\Pi}^{IDP-UNF} = Pr[Event] + Pr[\overline{Event}]$. Define $\epsilon_1 = Pr[Event]$ and $\epsilon_2 = Pr[\overline{Event}]$.

When $Event$ occurs, we construct the algorithm $\mathcal{B}$. It uses $\mathcal{A}$ as a subroutine, and solves the one-wayness of $gen(\cdot)$. Let ε be the NIZK extractor for $\mathcal{A}$. The algorithm $\mathcal{B}$ works as follows.

1. $\mathcal{B}$ randomly selects $i \in \{1, ..., n\}$.
2. $\mathcal{B}$ performs the experiment $Exp_{\mathcal{A},\Pi}^{IDP-UNF}(\lambda)$ with $\mathcal{A}$ to obtain $(pub^*, proof_{id}^*)$.
3. $\mathcal{B}$ runs the $\varepsilon(vk_{proof}, \pi_{proof}^*)$ to obtain $w_{proof} = \{path_{id}^*, CH_{id}^*, pk_{chash}^*, sk_{chash}^*, priv^*, r^*, rn^*\}$.
4. If $pub^* = pub_i$, then $\mathcal{B}$ outputs $priv^*$; otherwise, $\mathcal{B}$ aborts.

Because the index i is selected at random, $\mathcal{B}$ succeeds with probability ϵ_1/n. Because of the one-wayness of the $gen(\cdot)$, ϵ_1 must be negligible in λ.

When $\overline{Event}$ occurs, we construct algorithm $\mathcal{Z}$. It uses $\mathcal{A}$ as a subroutine and finds collision for the chameleon hash scheme. $\mathcal{Z}$ sends $(chashset, proof_{id})$ to O_{reg}, and obtains $P_{chash} = \{CH_{id_1}, ..., CH_{id_k}\}$ from the oracle O_{reg}, where $k \ll \lambda$. The set P_{chash} includes the chameleon hash CH_{id} of the user whose identity is id. The algorithm $\mathcal{Z}$ performs as follows.

1. $\mathcal{Z}$ randomly selects $i \in \{1, ..., k\}$.
2. $\mathcal{Z}$ performs the experiment $Exp_{\mathcal{A},\Pi}^{IDP-UNF}(\lambda)$ with $\mathcal{A}$ to obtain $(pub^*, proof_{id}^*)$.
3. $\mathcal{Z}$ runs the $\varepsilon(vk_{proof}, \pi_{proof}^*)$ to obtain $w_{proof} = \{path_{id}^*, CH_{id}^*, pk_{chash}^*, sk_{chash}^*, priv^*, r^*, rn^*\}$.
4. If $CH_{id}^* = CH_{id_i}$, then $\mathcal{Z}$ outputs $(priv^*, r^*)$; otherwise, $\mathcal{B}$ aborts.

Because the index i is selected at random, $\mathcal{Z}$ succeeds with probability ϵ_2/k. Furthermore, because of the collision resistance of the chameleon hash scheme, ϵ_2 must be negligible in λ.