ORIGINAL RESEARCH

Aitac: an identity-based traceable anonymous communication model

Fengyin Li[1] · Zhongxing Liu[1] · Yilei Wang[1] · Nan Wu[2] · Jiguo Yu[3,4] · Chongzhi Gao[5] · Huiyu Zhou[6]

Abstract
In the big data background, data privacy becomes more and more important when data leakage and other security events occur more frequently. As one of the key means of privacy protection, anonymous communication attracts large attention. Aiming at the problems such as low efficiency of message forwarding, high communication delay and abusing of anonymity, this paper presents an identity-based traceable anonymous communication model by adding a preprocessing phase, modifying the ciphertext structure and increasing the controllability of anonymity. Firstly, a new identity-based signature algorithm is proposed, and its security is proved via existential unforgeability against chosen-message attacks (EU-CMA). The signature algorithm is further applied to the anonymous communication model to implement the controllability of revocable anonym-ity. Secondly, by adding a preprocessing Setup phase, the operations of identifications distribution and user authentication are launched before the anonymous communication phase starts, and this practice significantly improves the efficiency of the anonymous communication model. Finally, by adding the hash value of the message and the user identification as the message authentication code, we design a new ciphertext structure, which can efficiently guarantee the integrity of the ciphertext. Performance analysis and simulation results show that the proposed anonymous communication model has high message forwarding efficiency and better security and controllability of anonymity.

Keywords Identity-based encryption · Bilinear map · Privacy protection · Identifications · Anonymous communications

1 Introduction

With the development of network technology and increasing big data applications, data privacy attracts more and more attention (Li et al. 2009; Jayaraman and Panneerselvam 2020; Silva et al. 2020; Wang et al. 2020). As one of the key means of privacy protection, anonymous communication has become a major research focus. Anonymous communication technology originated from the MIX-net mechanism (Chaum 1981), which confuses messages through a single or multiple MIX nodes to hide the user's identity. Later, two more anonymous communication schemes, TOR (The Onion Router) network (Dingledine et al. 2004; Hiller et al. 2019) and DC-net mechanism (Chaum 1988), were proposed based on the onion routing algorithm and standard cryptography technology respectively. Subsequently, anonymous communication has been developed rapidly in neural network (Li et al. 2018a), cloud computing (Li et al. 2018b) and Internet of Things (Corrigangibbs et al. 2013), and become an indispensable technology in the fields of information security and privacy protection. TOR networks achieve anonymity by layered message encryption in the public key cryptosystem. However, TOR cannot resist traffic analysis attacks (Bauer et al. 2007; Bai et al. 2008), and requires every entity to be entirely honest, especially the first and the last nodes. Otherwise, TOR network will be broken with the collusion of the malicious nodes. The DC-net mechanism obtains identity anonymity by arranging the

✉ Chongzhi Gao
czgao@gzhu.edu.cn

1 School of Computer Science, Qufu Normal University, Rizhao 276826, China

2 Science and Technology Department, Qufu Normal University, Qufu 273165, China

3 School of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China

4 Shandong Computer Science Center, National Supercomputer Center in Jinan, Jinan 250014, China

5 Institute of Artificial Intelligence and Blockchain, Guangzhou University, Guangzhou 510006, China

6 School of Informatics, University of Leicester, Leicester LE1 7RH, UK

entity to operate the same operation simultaneously in the given time period. This method can resist traffic analysis attacks effectively and assure the system security with the existence of a few dishonest nodes. However, the DC-net model requires all the users to be online at the same time conducting anonymous communication, which is unfeasible in applications (Hoang and Pishva 2014). To address DC-net's disadvantage of simultaneous online appearance, Jiang et al. (2018) proposed Acibe, an identity-based anonymous communication model, in 2018. The model allows the user to upload and download multiple messages in each time period, and encrypts messages through the identity-based encryption scheme (IBE). However, the anonymous communication efficiency dramatically decreases with the increase of the number of communication messages, and the message integrity cannot be assured in the Acibe model.

To address these problems, by adding the preprocessing Setup phase before anonymous communication, an identity-based traceable anonymous communication model is proposed in this paper. In our new model, users only download and decipher the ciphertexts belonging to themselves, thus the expense of anonymous communication is significantly reduced. Furthermore, the proposed new signature algorithm can assure the users to selectively reveal their identities to other members if necessary, implementing the traceability of anonymity in the anonymous communication model.

The main contributions of this paper are as follows:

1. We propose a new identity-based signature algorithm, and proves its security in the model of EU-CMA. Furthermore, it is applied to the anonymous communication model to implement the revocable anonymity, and enabling the controllability of anonymity in the anonymous communication model.
2. By adding a preprocessing Setup phase, we achieve preprocessing operations of identifications distribution and identity authentication between the users. Thus, the efficiency of the anonymous communication phase is largely improved.

   Furthermore, users can pick up ciphertexts sent to them by interpreting the identifications prefix before downloading. In this way, the decryption cost of anonymous communication is greatly reduced, and the efficiency of message forwarding is improved.
3. By adding the hash values of the messages and the identification as the message authentication code, this paper presents a new ciphertext structure to guarantee the integrity of ciphertexts. Thus, the receiver can verify the integrity of the decrypted messages, preventing the adversary from replacing the identifications, and improving the security of the anonymous communication model.

This paper is organized as follows. Section 2 introduces the basics. Section 3 presents the proposed digital signature scheme and justification, Sect. 4 reports the traceable anonymous communication model with performance analysis. Finally, Sect. 5 concludes this paper.

2 Preliminaries

2.1 Bilinear map

Bilinear mapping (Yu and Li 2019) is a function that generates elements in the third vector space from the elements in a two vector space. It can be described by a quaternion: (p, G_1, G_2, e). G_1, G_2 are two cyclic groups of prime order p, and e is a mapping $e : G_1 * G_1 \rightarrow G_2$. e satisfies the following properties:

1. Bilinear: For any $a, b \in Z_p, P, Q \in G_1$, there is $e(P^a, Q^b) = e(P, Q)^{ab}$.
2. Non-degeneracy: There is generator P in cyclic group G_1 that the following formula holds: $e(P, P) \neq 1_{G_2}$(1_{G_2} is the identity element of cyclic group G_2).
3. Computability: There is an effective algorithm that can compute the value of $e(P, Q)$ for any $P, Q \in G_1$.

2.2 Identity-Based Encryption

Identity-Based Encryption (IBE) is an encryption scheme with the core of the trusted key generation center (KGC) (Shamir 1985). The scheme generates the private key for the user by using the user's ID, and the user's the public key can be derived from the user's public information, such as ID and email address. In their scheme, the time for the user to pass the public key can be omitted, because each user knows the ID of the other user(s) and can encrypt messages with it. This greatly solves the storage problem of public key certificates. The scheme consists of four probabilistic polynomial time stages: (1) Initialization; (2) Key generation; (3) Encryption; (4) Decryption.

1. Initialization: KGC randomly selects public parameters k and system master key msk, and sets the plaintext space M and ciphertext space C.
2. Key generation: KGC generates private key d using the user's ID and master key msk, and passes it back to the user.

3. Encryption: The sender uses the public key ID of the receiver to encrypt the message $m \in M$ and obtains ciphertext $c \in C$.
4. Decryption: The receiver uses his private key d to decrypt the ciphertext c and obtains the message m.

2.3 Strong diffie-hellman problem assumption

In the bilinear mapping groups (G_1, G_2), g_1 and g_2 are the generators of G_1 and G_2 respectively, where G_1 may be equal to G_2. Then the q-SDH problem is defined as: Given a (q+2)-tuple as input, the adversary outputs a pair $(c, g_1^{\frac{1}{a+c}})$, where $c \in_R Z_p^*$. Then the advantage of the adversary in solving the q-SDH problem is defined as:

$$A_{Adv}^{SDH} = \Pr[A(g_1, g_2, g_2^a, \ldots, g_2^{a^q}) = (c, g_1^{\frac{1}{a+c}})] \geq \varepsilon.$$

The SDH problem is safe if and only if probability A_{Adv}^{SDH} of a successful attack by the adversary A in the polynomial time is negligible.

2.4 Symbolic meaning

In this section, we list the important symbols used in this paper and their definitions, as shown in Table 1.

3 A new identity-based signature scheme

Using the IBE and the Boneh-Boyen schemes (Boneh and Boyen 2004), this paper presents a new signature algorithm for plaintext $m \in Z_p^*$:

1. Initialization: Given the bilinear cyclic group parameter PG = (G_1, G_2, p, g, e, H), where G_1, G_2 are two cyclic groups of order p, g is the generator of G_1, $e : G_1 \times G_1 \to G_2$ is the bilinear mapping, and $H : Z_p^* \to G_1$ is the collision-resistant hash function.
2. Key generation: Given the user's ID, the key generation algorithm computes $y = H(ID)$, and then randomly selects $x \in Z_p^*$, computes $g_3 = g^x$, and returns a public/secret key pair (pk, sk) as follows:

   $$pk = (g_3, y), \quad sk = x.$$

3. Signature: The signer randomly selects $k \in Z_p^*$ and uses private key $sk = x$ to compute the signature $\sigma = (\sigma_1, \sigma_2) = (g^k, y^{\frac{1}{x+k+m}})$ for the message m, and returns (m, σ).
4. Verification: The receiver uses the received signature (m, σ) and the signer's public key $pk = (g_3, y)$ to verify the validity of the signature. The receiver accepts the signature if equation $e(\sigma_2, \sigma_1 g_3 g^m) = e(g, y)$ holds.
5. Proof of the correctness of the signature scheme:

$$\begin{aligned} e(\sigma_2, \sigma_1 g_3 g^m) &= e(y^{\frac{1}{x+k+m}}, g^k g_3 g^m) \\ &= e(y^{\frac{1}{x+k+m}}, g^{x+k+m}) = e(g, y) \end{aligned} \tag{1}$$

6. Proof of the security of the proposed signature scheme:

Theorem 1 *If the q-SDH problem is hard, the signature scheme is provably secure in the EU-CMA security model.*

Proof Suppose there exists an adversary A who can break the signature scheme with (t, q_s, ε) in the EU-CMA security model. We construct a simulator B to solve the q-SDH problem with the ability of the adversary A. Given a problem instance $(g, g^a, g^{a^2}, \ldots, g^{a^q})$ over the pairing group parameter PG, B and A cooperatively works as follows.

Setup Let $SP = PG$, B randomly chooses $\omega_0, \omega_1, \ldots, \omega_q \in Z_p^*$ and computes the public key $pk = (g_3, y)$ from the problem instance and the chosen parameters:

$$g_3 = g^a, y = g^{\omega_0(a+\omega_1)\cdots(a+\omega_q)},$$

where the secret key $x = a$, and we require q = q_s.

Query The adversary makes signature queries in this phase. For the i-th signature query on m_i, B enables $k_i = \omega_i - m_i$ and uses the problem instance $g, g^a, g^{a^2}, \ldots, g^{a^q}$ and the chosen parameters $\omega_0, \omega_1, \ldots, \omega_q$ to compute the signature:

$$\begin{aligned} \sigma_{m_i} = (\sigma_1, \sigma_2) &= (g^{\omega_i - m_i}, g^{\frac{\omega_0(a+\omega_1)\cdots(a+\omega_q)}{a+\omega_i - m_i + m_i}}) \\ &= (g^{\omega_i - m_i}, g^{\omega_0(a+\omega_1)\cdots(a+\omega_{i-1})(a+\omega_{i+1})\cdots(a+\omega_q)}) \end{aligned} \tag{2}$$

According to the signature definition and the simulation algorithm, the following equation holds.

$$\sigma_2 = y^{\frac{1}{x+k_i+m_i}} = g^{\frac{\omega_0(a+\omega_1)\cdots(a+\omega_q)}{a+\omega_i - m_i + m_i}} = g^{\omega_0(a+\omega_1)\cdots(a+\omega_{i-1})(a+\omega_{i+1})\cdots(a+\omega_q)} \tag{3}$$

Table 1 Symbol list

Symbols	Meaning
d_i	The privacy key of user U_i
id_i	The public key of user U_i
n_{ij}	The identification set by user U_i to user U_j
N	The number of members in an anonymous group
C	The ciphertext
M	The message

therefore, σ_{m_i} is a valid signature of m_i.

Forgery The adversary returns a forged signature on the challenge message m^* that has not been queried:

$$\sigma_{m^*} = (\sigma_1^*, \sigma_2^*) = (g^{k^*}, y^{\frac{1}{x+k^*+m^*}}) \tag{4}$$

If equation $m^* + k^* = m_i\beta + r_i$ holds for some queried signature of m_i, abort. Otherwise, let $c = m^* + k^*$, we have $c \neq \omega_i = m_i + k_i$ for all $i \in [1, q_s]$, and $\sigma_2^* = y^{\frac{1}{x+k^*+m^*}} = y^{\frac{1}{x+c}} = g^{\frac{1}{\omega_0(a+\omega_1)\cdots(a+\omega_q)}}$.

Let $\sigma_2^* = g^{f(a)+\frac{d}{a+c}}$, where f(a) is a (q − 1)-degree polynomial function, d is a nonzero integer. The simulator B computes the following equation.

$$(\frac{\sigma_2^*}{g^{f(a)}})^{\frac{1}{d}} = (\frac{g^{f(a)+\frac{d}{a+c}}}{g^{f(a)}})^{\frac{1}{d}} = g^{\frac{1}{a+c}} \tag{5}$$

and outputs $(c, g^{\frac{1}{a+c}})$ as the solution to the q-SDH problem instance.

Considering the hardness of the q-SDH problem, theorem 1 is proved.

Indistinguishable simulation The randomness of the simulation includes random numbers in the key generation and the signature generation. Let $y = g^{\omega_0(a+\omega_1)\cdots(a+\omega_q)} = g^{\gamma}$. There are

$$x, \gamma, k_1, k_2, \ldots, k_{q_s} = a, \omega_0(a+\omega_1)\cdots(a+\omega_q),$$
$$\omega_1 - m_1, \omega_2 - m_2, \ldots, \omega_{q_s} - m_{q_s}.$$

According to the setting of the simulation, where a, ω_0, ω_i are randomly chosen, we can see that they are random and independent from the point of view of the adversary. Therefore, the simulation is indistinguishable from the real attack.

Probability of successful simulation and useful attacks Suppose there is no abortion in the simulation. The random numbers in the queried signature and the forged signature are k_i and k^*, respectively. If $m^* + k^* \neq m_i\beta + r_i$ for all $i \in [1, q_s]$, then the forged signature can be reduced to the q-SDH problem. Because the probability at $m^* + k^* = m_i\beta + r_i$ is $1/q_H$, so the probability at $m^* + k^* \neq m_i\beta + r_i$ is $1 - 1/q_H$. Therefore, the probability of the successful simulation and attacks is $1 - 1/q_H$.

Advantage and time cost Suppose the adversary breaks the scheme with (t, q_s, ε). The advantage of solving the q-SDH problem is ε. Let T_s denote the time cost of the simulation. We have $T_s = O(q_s)$. Therefore, B will solve the q-SDH problem with $(t + T_s, \varepsilon)$.

4 Identity-based traceable anonymous communication model

In our proposed Aitac model, a preprocessing phase is added before the communication phase, so that each two members of the anonymous group has corresponding communication identifications, which reduce the number of ciphertexts downloaded in the communication phase and reduces the cost of anonymous communication. At the same time, by adding the message authentication code, the problem of identification replacement attack is solved. Finally, the previous signature scheme is applied to the Aitac model so that the users can choose to remove the anonymity if necessary. Our model is mainly composed of five parts: the overall architecture, preprocessing phase, anonymous communication phase and performance analysis of the model.

4.1 The architecture

4.1.1 Time scheduling scheme

By adding the Setup phase to the time scheduling of the Acibe model, the time scheduling scheme for our proposed Aitac model is shown in Fig. 1.

The model proposed in this paper is composed of two phases: Setup and anonymous communication phases. The Setup phase is organized into five time periods: T_1, T_2, T_3, T_4, T_5. During period T_1, KGC generates the user's private key d and passes it back to the user. During period T_2, each user generates corresponding identifications for other $N-1$ members, and uses the corresponding user's

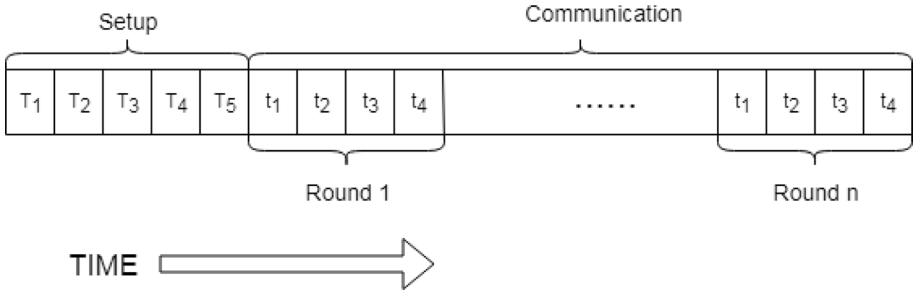

Fig. 1 Time scheduling scheme

public key *id* for encryption to obtain the ciphertext. During period T_3, each user needs to upload $N-1$ ciphertext encrypted and sent to the bulletin board. During period T_4, each user in the anonymous group needs to download all the ciphertexts on the bulletin board. During period T_5, each user tries to decrypt all the downloaded ciphertexts using his/her own private key. If the ciphertext is encrypted with his/her own public key, the corresponding identification n and the ID of the sender can be obtained from the decryption result.

The anonymous communication phase is organized into several cycles, and each cycle is organized into four periods: t_1, t_2, t_3, t_4. During period t_1, the sender uses the public key *id* of the receiver to encrypt the plaintext and retrieve the corresponding ciphertext. During period t_2, each user uploads the ciphertext to the bulletin board at least once. During period t_3, each user identifies the prefixes (namely identifications n) of all the ciphertexts on the bulletin board, and then downloads the ciphertexts belonging to their own identifications. At the same time, it is necessary to ensure that each user downloads the ciphertext at least once for updating during this period. If there is no corresponding ciphertext on the bulletin board, any ciphertext on the bulletin board will be randomly downloaded. During period t_4, the receiver uses his/her private key d to decrypt the downloaded ciphertexts and recovers the plaintexts.

Specific processes of each phase will be explained in the following Sects. 4.2 and 4.3.

4.1.2 Security goal

The security of the anonymous communication model is reflected in the following aspects:

1. Confidentiality of message $_m$: As a model using the IBE scheme, ensuring the confidentiality of the plaintext is a basic requirement. In this paper, the public key system is used for encryption, so that the confidentiality of the plaintext is firmly guaranteed in the proposed Aitac model.
2. Anonymity of the sender: In the anonymous group, except the sender and the receiver, other $N-2$ members do not know the identity of the uploaders corresponding to the ciphertext on the bulletin board. The senders can also selectively use the signature algorithm to generate their own signatures to ensure that other users know their true identities.
3. Anonymity of the receiver: Other members of the anonymous group do not know the identity of the receiver, and only the sender knows the identity of the receiver in this communication.

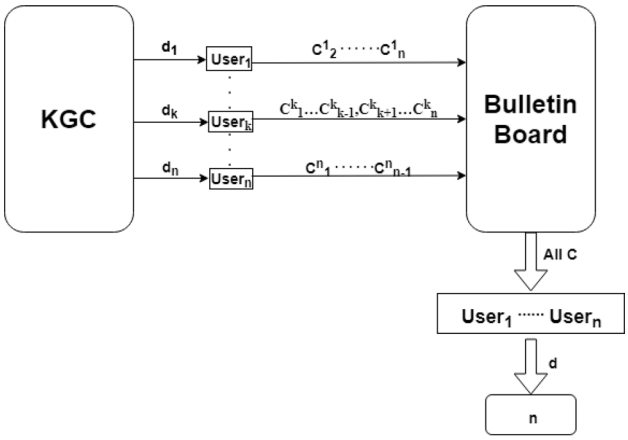

Fig. 2 The processes of the Setup phase

4.2 Preprocessing phase

This phase can also be called the Setup phase, which mainly generates and transmits the user's private key d and identifications n. The specific processes are shown in Fig. 2.

For the Setup phase, this paper is organized into three parts: (1) The generation of public parameters; (2) The generation of the user's private key; (3) Generation and transmission of identification n.

1. The generation of public parameters: Let G_1, G_2 be two cyclic groups of order p, and there is a bilinear mapping $e : G_1 \times G_1 \to G_2$. Suppose g is the generator of cycle group G_1. KGC randomly selects $\lambda \in Z_p^*$ as the system master key, and let $g_1 = g^{\lambda}$. g_2 is randomly selected from G_1. The public parameters are $PG = (G_1, G_2, g, g_1, g_2, e, p)$.
2. The generation of the user's private key: KGC randomly selects $r \in Z_p^*$, then computes the user's private key $d = (d_1, d_2, d_3,) = (g_2^{\lambda} g_1^{id \cdot r}, g^r, g^{id \cdot r})$ according to the user's identity $id(id \in Z_p^*)$, and passes the private key back to the user. The user's *id* is the public key.
3. Generation and transmission of identifications n_{ij}:

At this step, each user $U_i(i = 1, 2, \cdots, N)$ in the anonymous group generates $N-1$ large random number $n_{ij} \in G_2(j = 1, 2, \cdots, N, j \neq i)$, which acts as the identifications label for the remaining $N-1$ users $U_j(j = 1, 2, \cdots, N, j \neq i)$ in the following anonymous communication.

At the same time, user U_i randomly generates parameters $t_{ij}, s_{ij} \in Z_p^*(i, j = 1, 2, \cdots, N, j \neq i)$, and uses the

corresponding public keys $id_j (j = 1, 2, \cdots, N, j \neq i)$ of user U_j to encrypt his/her identification. Then,U_i puts the key hash value $h^{id_j}(id_i \parallel n_{ij})$ of the identity id_i and the identification n_{ij} in the position of the second parameter as the identity authentication code. Finally, the identity of the sender is added as the first parameter of the ciphertext. The format of the ciphertext is:

$$c_i' = (c_0', c_1', c_2', c_3', c_4', c_5') = (id_i, h^{id_j}(id_i, n_{ij}), e(g_1, g_2)^{t_{ij}} \cdot n_{ij}, g_1^{id_j(t_{ij}+s_{ij})}, g_1^{s_{ij}}, g^{t_{ij}}) \tag{6}$$

Then the sender U_i uploads the encrypted ciphertext c_i' to the bulletin board. As shown in Fig. 2, $C_2^1, \ldots, C_n^1$ is the ciphertext of the identifications generated by the first user for the other *N-1* users.

After the upload step, each user in the anonymous group must download all the ciphertexts on the bulletin board and then decrypt all the ciphertexts according to their own private keys d_j=$(d^1, d^2, d^3) = (g_2^{\lambda} g_1^{id_j \cdot r}, g^r, g^{id_j \cdot r})$,

$$n_{ij} = c_2' \frac{e(d^2, c_3')}{e(d^1, c_5')e(d^3, c_4')} = \frac{n_{ij} \cdot e(g_1, g_2)^{t_{ij}} e(g^r, g_1^{id_j(t_{ij}+s_{ij})})}{e(g_2^{\lambda} g_1^{id_j \cdot r}, g^{t_{ij}})e(g^{id_j \cdot r}, g_1^{s_{ij}})} \tag{7}$$

If it is a ciphertext encrypted with the receiver's own public key, it can be decrypted to get the ciphertext n_{ij}. At the same time, the receiver can hash the decrypted identification n_{ij} with the first parameter, that is, the identity id_i of the sender, to examine whether or not the hash value obtained is consistent with the second parameter $h^{id_j}(id_i, n_{ij})$ in the ciphertext. If consistent, the plaintext is proved to be correct; Otherwise, the identification n_{ij} is wrong or the identity id_i is replaced by the attacker.

Note: Each user is required to upload at least once during the transmission of identifications and download all the ciphertexts.

4.3 Anonymous communication phase

4.3.1 Anonymous communication model

After each user retrieves his/her own selected $N-1$ identifications from the other $N-1$ users during the Setup phase, he/she can filter out his/her own ciphertexts from the bulletin board before downloading them during the communication phase. That is, he/she does not need downloading and decrypting other ciphertexts unrelated to him/her. At the same time, he/she knows the identity of the sender from the prefix of the ciphertext. This filtering operation greatly reduces the burden of downloading and decrypting the relevant operations in the anonymous communication phase. The specific communication process is shown in Fig. 3.

The anonymous communication phase is composed of four steps:

1. Encryption:
   For the given message m , the sender U_i firstly computes the ciphertext by the identity-based encryption algorithm. Then U_i assigns the identificationn_{ij} of the corresponding receiver U_j as a prefix of the ciphertext, and uses the keyed hash value $h^{id_j}(m \parallel n_{ij})$ of message m and identification n_{ij} in the position of the second parameter as the message authentication code. The receiver's public key id_j is used as the key of the hash function. We can retain the following ciphertext structure:

$$\begin{aligned} c &= (c_0, c_1, c_2, c_3, c_4, c_5) \\ &= (n_{ij}, h^{id_j}(m \parallel n_{ij}), e(g_1, g_2)^{t_{ij}} \cdot m, g_1^{id_j(t_{ij}+s_{ij})}, g_1^{s_{ij}}, g^{t_{ij}}) \end{aligned} \tag{8}$$

   As shown in Fig. 3, $C_1 = Enc(m_1)$ indicates that the sender U_1 encrypts the message m_1 by performing the above encryption formula to obtain the ciphertext:

$$Enc(m_1) = (n_{1j}, h^{id_j}(m_1 \parallel n_{1j}), e(g_1, g_2)^{t_{1j}} \cdot m_1, g_1^{id_j(t_{1j}+s_{1j})}, g_1^{s_{1j}}, g^{t_{1j}}) \tag{9}$$

   where the first parameter of the ciphertext is the identification n_{1j}, which indicates that the receiver of the ciphertext is U_j and its public key is id_j.
2. The sender uploads the ciphertext: In the upload step, each user uploads the ciphertext at least once;
3. The receiver downloads the ciphertext on the bulletin board: In the download step, the receiver filters the ciphertext according to the prefixes (namely, the identifications n_{ij}) of the ciphertext, and selects the ciphertext belonging to him/herself to download. C_i, C_j, C_h, described in Fig. 4, are the ciphertexts with the identifications by each receiver ($User_i, User_j, User_h$) filtering

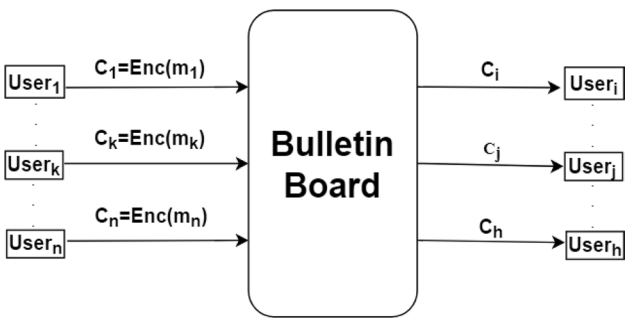

Fig. 3 The specific process of the communication phase

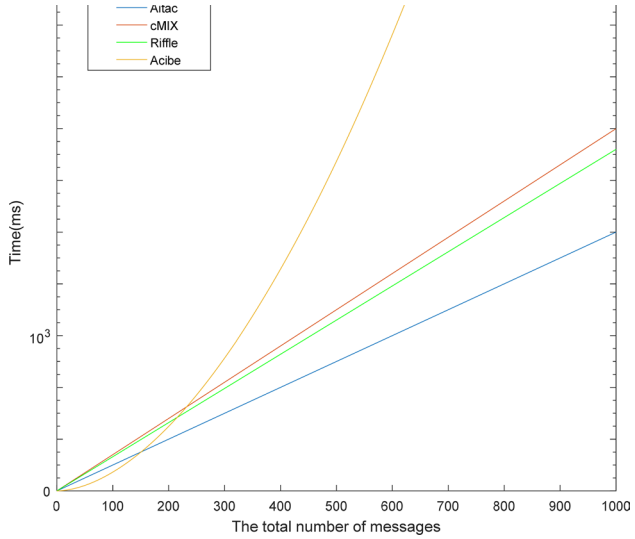

Fig. 4 Comparison of time cost in decryption step

the irrelevant information on the bulletin board. At this phase, each user is required to download at least once. If the receiver retrieves all the ciphertext and finds that there is no identification belonging to him/herself, and downloads the first ciphertext on the bulletin board.

4. The receiver decrypts the downloaded ciphertext: The last step of the communication phase is that the receiver decrypts the ciphertext downloaded in the third step with his/her private key d_j=$(d^1, d^2, d^3) = (g_2^{\lambda} g_1^{id_j \cdot r}, g^r, g^{id_j \cdot r})$ to obtain the clear message m:

$$m = c_2 \frac{e(d^2, c_3)}{e(d^1, c_5) e(d^3, c_4)} \tag{10}$$

At the same time, the receiver can hash the decrypted message m with the first parameter, that is, the identification n_{ij}, to evaluate whether or not the hash value obtained is consistent with the second parameter $h^{id_j}(m \parallel n_{ij})$ in the ciphertext. If consistent, the plaintext is proved to be correct; Otherwise, the message m is wrong or the identification n_{ij} is replaced by the attacker's.

4.3.2 Traceability of the proposed Aitac Scheme

The signature algorithm we designed in Sect. 3 can be applied to the anonymous communication model in this paper, that is, the users can use the signature algorithm to generate their own signatures, selectively remove the anonymity in the communication phase, and realize the traceability of anonymity.

Since message m belongs to Z_p^* in the signature scheme, we first define a collision-resistant hash function: $H' : G_2 \to Z_p^*$, then compute $h' = H'(m)$, and then apply the above signature algorithm so that all the users in the anonymous group know the real identity of the sender, while ensuring the privacy of message m:

1. Signed key generation: Given the user's id, the key generation algorithm computes $y' = H'(id)$, and then randomly selects $x' \in Z_p^*$ as the signed privacy key, and $(g^{x'}, y')$ as the signed public key.
2. Signature: The sender randomly selects $k' \in Z_p^*$ and uses the private key x' to compute the signature $\sigma_{sender_{id}} = (\sigma'_1, \sigma'_2) = (g^{k'}, y'^{(\frac{1}{x'+k'+h'})})$ for h', and returns $(h', \sigma_{sender_{id}})$.

Then the sender adds the ciphertext $c = (c_1, c_2, c_3, c_4) = (e(g_1, g_2)^{t_{ij}} \cdot m, g_1^{id_j(t_{ij}+s_{ij})}, g_1^{s_{ij}}, g^{t_{ij}})$ and the hash value h' used in the signing process of the ciphertext structure to be sent, and the format of the ciphertext with the additional signature is:

$$C = (C_0, C_1, C_2, C_3, C_4, C_5) = (n_{ij}, h^{id_j}(m \parallel n_{ij}), \sigma_{sender_{id}}, id_i, c, h') \tag{11}$$

If the ciphertext sent by the sender contains signature information, all the users know the sender's identity in the ciphertext during the communication phase, and then signature $\sigma_{sender_{id}}$ can be verified using the sender's signature public key pk. Through verification, the result can be compared with the last parameter h' in the ciphertext. If it is consistent, the identity of the sender is id_i. In this way, the anonymity of the communication phase can be removed, and the message m can be obtained by decrypting c with the sender's encrypted public key.

4.4 Performance analysis

4.4.1 Theoretical analysis

This paper mainly investigates the performance of the proposed model from three aspects: traceability, security and the efficiency of communication.

1. Traceability: This paper achieves the traceability of the anonymity of the communication model through the signature scheme proposed above. The sender generates his/her own signature through the signature scheme and append the signature to the third parameter position of the ciphertext structure, that is $C = (n, h^{ID}(m \parallel n), sign_{sender_{id}}, id, c, h')$. By publishing the sender's id, all the users can know and verify the sender's identity. In this way, the anonymity of the communication model is removed.
2. Security: Security analysis is conducted according to the three security goals proposed in the model, namely,

Table 2 The comparison of the security by different models

Targets	DC-net	Acibe	Aitac
Security of the message	√	√	√
Anonymity of the sender	√	√	√
Anonymity of the receiver	√	√	√
Traceability of anonymity	×	×	√

Table 3 The time costs of each phase by different models

Targets	DC-net	Acibe	Aitac
The time cost of encryption process	O(MN)	O(M)	O(M)
The time cost of upload process	O(MN)	O(M)	O(M)
The time cost of download process	O(MN)	O(MN)	O(N)
The time cost of decryption process	O(MN)	O(MN)	O(N)

Note: M represents the number of the messages, and N represents the number of the users in the anonymous group

the security of the message, the anonymity of the sender and the anonymity of the receiver. The security of the message can be guaranteed by the identity-based encryption (IBE) scheme. As long as the encryption scheme is secure, the message is secure. We here adopt the IBE scheme proposed in the ElGamal public-key system (Elgamal 1984), which can protect the anonymity and resist the attacks of any CPA adversary. The anonymity of the sender is guaranteed by the central storage structure of the bulletin board. All the senders' messages are uploaded to the bulletin board for centralized obfuscation, so it is impossible to know which message corresponds to the sender. The anonymity of the receiver is guaranteed by two parts: the IBE scheme and the downloading phase of communication. First, the IBE scheme ensures that when the sender encrypts the receiver's public key ID, other users do not know the receiver's ID. In the downloading phase of communication, each user is required to download the ciphertext at least once, which can ensure that there is no case such as "only one user downloads" so the receiver's anonymity can be guaranteed.

3. The efficiency of communication: First, the model adds the Setup phase before communication, which seems to be more complicated, but by joining the Setup phase, each two users in the anonymous group have a unique identification. Therefore, the user can filter the ciphertext during the downloading phase, which greatly reduces the download volume and improves the efficiency of the communication process. The Setup phase only takes up the first period in the entire communication period, and all the rest of the communication period allows us to filter the ciphertext with the identification, thus the efficiency of communication can be secured. In general, the efficiency of the entire model is improved significantly.

For the description of the above three aspects, we compare the proposed model with the Acibe and DC-net models. The comparison of the security for the three models is shown in Table 2.

At the same time, according to the time costs of each phase in the three models, the comparison is shown in Table 3.

4.4.2 Experiment simulation

The model in this paper is implemented in Java, where the bilinear mapping is written using jPBC library in Java. The simulation environment of the whole model is built on a PC with a CPU 2.13 GHz and 6 GB of RAM. The performance of the model is compared with that of Acibe, cMIX (Chaum et al. 2017) and Riffle (Kwon et al. 2015). Because the model proposed in this paper has the Setup phase, each user in the anonymous group generates identifications for the other $N-1$ members, and the ciphertext of identifications by encrypting their passes to corresponding users so that the identification can be used as a ciphertext prefix. Thus, the users can collect the ciphertext information by identifying the prefixes without downloading all the ciphertexts. As less ciphertexts are downloaded, we have fewer decryption operations, and the time cost of the decryption step is much lower. This paper simulates a complete communication process, ignoring the communication cost in the upload and download phases, and gradually increases the number of the messages to 1000 in one communication cycle. We observe the changes of communication time with the increase of message numbers, and get the time cost in the decryption step as shown in Fig. 4.

In the private key generation phase, the Acibe model requires that the user's private key be recalculated every communication cycle, while the model proposed in this paper only requires the private key to be calculated once during the Setup phase, and all the subsequent communication cycles use the calculated key. Therefore, the time cost of the private key generation in our model is only related to the number of the users in the anonymous group, but not to the number of the messages that need to be communicated. This greatly simplifies the calculation of private key generation and improves the efficiency of anonymous communication. The time cost comparison of the private key generation is shown in Fig. 5.

Regarding the encryption of messages, the model described in this paper is the same as the Acibe model, so the time cost of the encryption process increases logarithmically with the increase in the number of messages. In

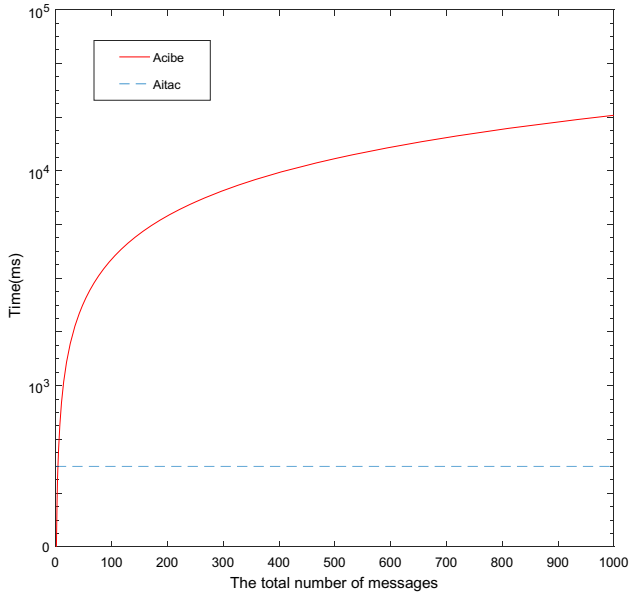

Fig. 5 The time cost comparison of the private key generation

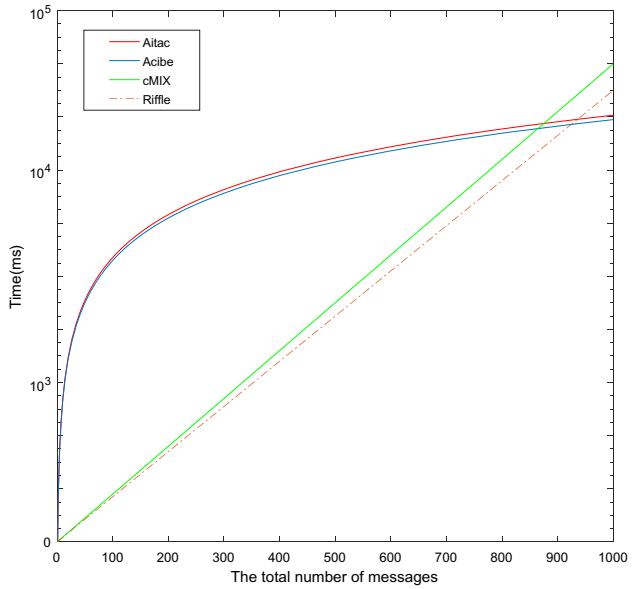

Fig. 6 The time cost comparison of the encryption

the cMIX and Riffle models, the time cost of the encryption with the increase of the number of messages, there is a linear growth. Therefore, the Riffle and cMIX have good encryption performance when the number of messages is few, but with the increase of the number of members and communication messages, the time cost has increased significantly. The time cost comparison of the encryption is shown in Fig. 6.

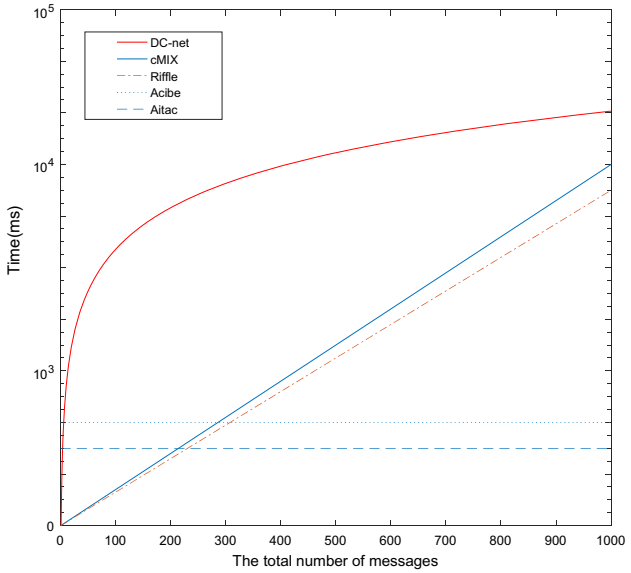

Fig. 7 Comparison of time cost in the entire communication period

The model in this paper compares with the DC-net model in the whole communication phase, as the number of messages increases, the time cost decreases significantly. Even compared with the Acibe model, there is no limit on the number of messages in a communication cycle, and the time cost of communication is still significantly reduced. In order to ensure strong anonymity, the Riffle and the cMIX model sacrifice the efficiency of communication, so the communication cost increases linearly with the increase in the number of messages. Therefore, the efficiency of communication in this paper is significantly improved compared with the previous communication models. The specifications are shown in Fig. 7.

Through the experimental analysis, the efficiency of the model proposed in this paper is mainly improved in the two phases: private key generation and decryption. Especially for a large number of anonymous messages that require frequent communication, the performance of the proposed model has been significantly improved.

5 Conclusions

In the big data area, data leakage and other security events occur frequently, and data privacy becomes more and more important. As one of the key solutions to protect data privacy, anonymous communication attracts research attention. To solve the problems of low efficiency during message forwarding, high delay of communication and abuse of anonymity in anonymous communication systems, this paper has presented an identity-based traceable anonymous

communication model by adding preprocessing operations, modifying the ciphertext structure and increasing the traceability of anonymity. Performance analysis and results of simulation show that the anonymous communication model proposed in this paper has better efficiency of message forwarding, better security and the traceability of anonymity. The decentralized edition of the proposed scheme will be implemented in the future work for further reducing computational complexity.

Acknowledgements This study was funded by Foundation of National Natural Science Foundation of China (Grant Numbers: 62072273, 61771231), the Major Basic Research Project of Natural Science Foundation of Shandong Province of China (ZR2018ZC0438), Natural Science Shandong Province (Grant Numbers: ZR2016FM23, ZR2017MF010, ZR2017MF062), Key Research and Development Program of Shandong Province (NO. 2019GGX101025).

References

Boneh D, Boyen X (2004) Short signatures without random oracles. Theory and application of cryptographic techniques, pp 56–73

Bai X, Zhang Y, Niu X (2008) Traffic identification of tor and web-mix. In: Proceedings of 8th IEEE International Conference on Intelligent Systems Design and Applications, pp 548–551

Bauer K, Mccoy D, Grunwald D et al (2007) Low-resource routing attacks against tor. Workshop on privacy in the electronic society, pp 11–20

Chaum D (1981) Untraceable electronic mail, return addresses, and digital pseudonyms. Commun ACM 24(2):84–90

Chaum D (1988) The dining cryptographers problem: unconditional sender and recipient intractability. J Cryptol 1(1):65–75

Chaum D, Das D, Javani F et al (2017) cMix: mixing with minimal real-time asymmetric cryptographic operations. In: International conference on applied cryptography and network security, pp 557–578

Corrigangibbs H, Wolinsky DI, Ford B, et al (2013) Proactively accountable anonymous messaging in verdict. In: usenix security symposium, pp 147–162

Dingledine R, Mathewson N, Syverson P (2004) Tor: The second-generation onion router. J Franklin Inst 239(2):135–139

Elgamal T (1984) A public key cryptosystem and a signature scheme based on discrete logarithms. IEEE Trans Inf Theory 31(4):469–472

Hiller J, Pennekamp J, Dahlmanns M et al (2019) Tailoring onion routing to the internet of things: security and privacy in untrusted environments. In: International conference on network protocols, pp 1–12

Hoang NP, Pishva D (2014) Anonymous communication and its importance in social networking. In: The 16th IEEE International Conference on Advanced Communication Technology, pp 34–39

Jayaraman I, Panneerselvam AS (2020) A novel privacy preserving digital forensic readiness provable data possession technique for health care data in cloud. J Ambient Intell Human Comput. https://doi.org/10.1007/s12652-020-01931-1

Jiang L, Li T, Li X et al (2018) Anonymous communication via anonymous identity-based encryption and its application in IoT. Wirel Commun Mobile Comput. https://doi.org/10.1155/2018/6809796

Kwon A, Lazar D, Devadas S et al (2015) Riffle: an efficient communication system with strong anonymity. Privacy Enhanc Technol 2016(2):115–134

Li F, Ma J, Li J (2009) Distributed anonymous data perturbation method for privacy-preserving data mining. J Zhejiang Univ Sci A 10:952–963

Li Y, Wang G, Nie L (2018a) Distance metric optimization driven convolutional neural network for age invariant face recognition. Pattern Recogn 75:51–62

Li J, Zhang Y, Chen X et al (2018b) Secure attribute-based data sharing for resource-limited users in cloud computing. Comput Secur 72:1–12

Shamir A (1985) Identity-based cryptosystems and signature schemes. Lect Notes Comput 196(2):47–53

Silva P, Casaleiro R, Simões P et al (2020) Risk management and privacy violation detection in the PoSeID-on data privacy platform. SN Comput Sci 1:188

Wang X, Xu Z, Cai Z et al (2020) Novel temporal perturbation-based privacy-preserving mechanism for smart meters. Mobile Netw Appl 25:1548–1562

Yu B, Li H (2019) Anonymous authentication key agreement scheme with pairing-based cryptography for home-based multi-sensor Internet of Things. Int J Distrib Sensor Netw. https://doi.org/10.1177/1550147719879379