Content-centric Group User Authentication for Secure Social Networks

Jian Shen, Member, IEEE, Anxi Wang, Chen Wang, Jiguoo Li, and Yan Zhang, Member, IEEE

Abstract—Content-centric computing systems, which can obtain users’ attributes by analyzing information from content, have been widely used in many fields. In social networks, the computing resources of users are restricted. Group authentication schemes are required to ensure the validity of the user and the accuracy of the shared data. In this paper, we propose a content-centric group user authentication scheme to guarantee the security and accuracy of shared data. The proposed group user authentication scheme makes use of the user’s content to generate the user’s feature vector. Furthermore, the group authentication scheme based on the identity of each user can guarantee the network security before data sharing. In addition, regular operations in the network are not affected or damaged by incidents that occur during the authentication process. The security and performance analyses show that our scheme is secure against various attacks and has lower computational cost than the existing schemes.

Index Terms—Security and privacy, user authentication, content-centric, social networks.

1 INTRODUCTION

Currently, the contents of networks are usually managed by servers or websites. Ordinary users can access data through end-to-end connections. Cloud-based networks, which were developed from client-server networks and peer-to-peer networks, are a type of content-centric network. In other words, users do not care where or how to obtain data but pay more attention to the meaning of the data itself. With the development of science and technology, content-centric computing systems, including data-oriented network architecture (DONA) [1], publish/subscribe Internet routing paradigm (PSIRP) [2], network of information (NetInf) [3], and content-centric networking (CCN) [4], have been considered by many researchers. In this paper, we mainly discuss how to describe social networks with content-centric computing systems [5]. Social networks are thought to be a type of content-centric network that share user content; that is to say, the communication between users in social networks is based on data rather than an IP address. A social network is characterized by the exchange of interests and content objects.

In social networks, shared information is related to the identity of users. When these shared data are stolen by attackers, the privacy of the data owners can be exposed. Therefore, methods to guarantee the legitimacy of the users have become a hot-spot in social networks [6]. In addition, users in social networks are usually organized into groups to share data [7]. Only users in the same group can gain access to the shared data. In a social network group, it is very important for the group manager to authenticate the legitimacy of the identities of users who want to join the group. Each user should be authenticated to guarantee the security of group data sharing [8, 9]. It is obvious that a group user authentication scheme is suitable for social networks to simultaneously share data and protect privacy.

The current user authentication schemes in social networks can be divided into the following two types: one-to-one authentication and group authentication. The authentication schemes mentioned in [10, 11] are one-to-one, where the certifier interacts with the verifier to verify the identity of the certifier [12]. For example, the RSA digital signature [13] can be used to authenticate the signer’s identity. In this approach, the verifier sends a random challenge to the signer. The signer then digitally signs the random challenge and returns the challenged digital signature to the verifier. If the returned signature is lugged, the verifier is sure that the signer is a user who has the identity of the public key digital certificate, which is used to verify the digital signature. If the scheme is applied to a social network, each user needs to be authenticated by the other users of the group, which can seriously increase the network communication overhead. For group authentication schemes in social networks [14, 15], a trusted manager that can participate in the entire certification process is necessary. In this type of authentication scheme, the trusted manager takes responsibility for information distribution and user registration, which reduces the clients’ computing and storage costs.

In fact, a social network has undefined boundaries, and data sharing is bound to a small group. In general, there are many different groups in social networks. Each group has different sharing topics, such as tourism and food. Members of a group have similar hobbies [16]. Users in a group are more concerned about the information itself than the source of the information [17]. In addition, it is necessary for network managers to remove malicious users to ensure normal operation. However, traditional authentication schemes do not take the combination of identifiers and interests into consideration. Therefore, in this paper, we present a content-centric group authentication scheme.
for social network groups by employing feature vectors of users’ comments, which enables multiple users to freely share data with high security and efficiency. Note that the pseudo-identity of each user is added to compute the user’s "identity". Moreover, the pseudo-identity is extracted from the user’s previous comments and user’s real ID.

1.1 Our Contributions

In this paper, we propose an efficient group authentication scheme with feature vector for user authentication in social networks that performs better than the state-of-the-art methods. Group authentication is presented to achieve mutual trust between users. Moreover, the network manager has no secret information about the users. In addition, the regular operations in networks are not affected or damaged by incidents that occur during the authentication process. The main contributions of this paper are summarized as follows.

- **Group user authentication scheme is proposed for social networks.** In this paper, a group authentication is achieved based on the user’s comments or shared data, which can be used to guarantee the security of communication among group users. The presented scheme can tolerate some malicious users. The authentication progress is completed among all legal users without failure. Moreover, in the authentication phase, the regular operation of the network is not affected or damaged by incidents that occur during the authentication process, which makes group data sharing in social networks more secure and reliable.

- **Content-centric computing system is used in social networks to provide users’ pseudo-identities.** According to the data sharing model, group users can form a group to efficiently share information. Note that the authentication phase is produced by the group user’s feature vector and the user’s real identity. A user’s feature vector is extracted from the user’s data by word frequency feature extraction. Attackers have no access to the identities of users generated from communication; thus, they cannot access the original comments. Therefore, the proposed group authentication scheme supports secure group data sharing in social networks.

- **The proposed group user authentication scheme is secure and efficient.** According to the security analysis, our scheme satisfies the security requirements of social networks and can resist a large number of attacks. In addition, the computational costs of users in one group are lower than those in the existing schemes. Moreover, we analyze the time cost of the proposed scheme using PBC library, which has been widely used to analyze group user authentication schemes.

1.2 Related work

In social groups, each user is only aware of its neighbors and has no knowledge about the information of other users. The group user authentication scheme has been widely used to protect the privacy of users and the security of shared data. However, Harn’s scheme only tolerates less than $t$ attackers which is decided by the polynomial function. After that, improved secret sharing schemes have been proposed in [28] and [29], which support less than $n/4 - 1$ attackers and less than $n/3$ attackers, respectively. In 2015, Miao et al. [15] found that the shared data in Harn’s scheme is not bound to each other and introduced the notion of a randomized component (RC) to bind share data with all users and to protect the share data from being exposed to the outside without computational assumptions. However, the new scheme can still only resist up to $t - 1$ colluded inside adversaries. Several researchers have designed new schemes [14, 30–32] to overcome the upper limit of $t$. However, the upper limit still exists in secret sharing schemes. In [18], Jiang et al. proposed an ECC-based authentication scheme which takes advantage of a temporal proof to avoid malicious tracking. However, the user’s secret information can be calculated by the gateway node through the shared old password. In [33], Shen et al. proposed a lightweight multi-layer authentication scheme. However, the secret information only stored in nodes will cause heavy security problems in the whole system. Note that Horng et al. [34] proposed a group communication scheme to allow entities to authenticate and securely communicate with each other in the group. This scheme does not have the upper limit.
of $t$. However, the invalid signature can severely affect the performance and the computational cost of each entity is a burden.

In contrast to the above solutions, in this paper, we design an efficient group authentication scheme by employing a trusted group manager (TGM) to collect the information of the users. The proposed scheme can achieve mutual trust between the users and the TGM, and a trusted cloud server (TCS) is used to generate random numbers for the users and the TGM. Moreover, the feature vector information is able to obtain the user’s secret information more quickly, and the user’s real identity can be protected.

1.3 Organization

The rest of the paper is sketched as follows. Section 2 introduces preliminary works for a better understanding of the topic. Section 3 presents the system model of the proposed scheme and discusses how to generate the users’ pseudo-identities. In Section 4, we introduce the details of the group user authentication scheme. In Section 5, we provide security analysis and comparison. In Section 6, our scheme is simulated by using the pairing-based cryptography (PBC) library. Then, we compare the performance with the existing related scheme. Finally, we conclude the paper in Section 7.

2 PRELIMINARIES

2.1 Cryptographic Bilinear Maps

Weil pairing [35] has been widely used in cryptography and public key cryptography. Here, we give an example to construct this map. Let $p$ be a prime such that $p = 6q - 1$ for some prime $q$ and $E$ be a supersingular elliptic curve defined by the Weierstrass equation $y^2 = x^3 + 1$ over $F_p$. The group of rational points $E(F_p) = \{(x, y) \in F_p \times F_p : (x, y) \in E\}$ forms a cyclic group of order $p + 1$. Furthermore, because $p + 1 = 6q$ for some prime $q$, the group of points of order $q$ in $E(F_p)$ forms a cyclic subgroup, denoted $G_1$. Further discussion of the Weil pairing is presented in the literature [36].

Definition 1. Let $G$ be a generator of $G_1$ and let $G_2$ be the subgroup of $F_p^*$ containing all elements of order $q$. A modified Weil pairing is a map $\tilde{e} : G_1 \times G_1 \rightarrow G_2$ that has the following properties for points in $E(F_p)$:

1. Bilinear: For any $P, Q \in G_1$ and $a, b \in Z$

   $\tilde{e}(aP, bQ) = \tilde{e}(P, Q)^{ab}$.

2. Non-degenerate: If $P$ is a generator of $G_1$, then $\tilde{e}(P, P) \in G_1$ is a generator of $G_2$. In other words, $\tilde{e}(P, P) \neq 1$.

3. Non-commutative: For any $P, Q \in G_1$, $P \neq Q$, $\tilde{e}(P, Q) \neq \tilde{e}(Q, P)$.

4. Computable: Given $P, Q \in G_1$, there exists an efficient algorithm to compute $e(P, Q)$.

2.2 Security Assumption

Security is one of the most essential conditions that a good cryptographic algorithm or scheme should first meet. Studies on safety issues boil down to the security model. The attacker’s ability and the goal of security achieved can be reflected by a correct and appropriate security model.

In this paper, we use the security model defined in the literature [37]. Note that the security of our scheme relies on the elliptic curve discrete logarithm problem (ECDLP). According to the proof in [38], the presented scheme can resist both passive attacks and active attacks. Many formal security analyses of the key agreement scheme can be found in the literature [39].

The ECDLP can be stated as follows. $P$ and $Q$ are two points on ECC. $k$ is an integer on $F_p$, such that $Q = kP$. For a given $P$ and $Q$, it is difficult to obtain $k$. If $k$ is sufficiently large, $k$ is the discrete logarithm of $Q$ to $P$ [40].

3 PROBLEM STATEMENT

3.1 System model

The whole model contains three primary entities: TCS, TGM and users. The relationship among the three entities is shown in Fig. 1. The TCS is an entity that provides cloud service [41]. In our model, the TCS is responsible for storing user data. The users, which can be an individual person or a major company, are entities that have similar hobbies in the same social group. The devices used to communicate with the TGM can include phones, computers, and tablets. The final component is the TGM, which is the group manager for the users. Generally, the TGM is equipped with professional knowledge of identity authentication so that it can provide convincing authentication results.

3.2 Feature vector

In a mixed social network, the characteristics of the user identities are very important. Similar to Facebook and Google+, it is difficult to directly extract the characteristics of a user by considering the user’s privacy. First, the connectivity degree of users in the network is relatively low owing to a restriction on the user’s contacts circle. In addition, there are a large number of visitors who have no characteristics in the network. Second, the user’s properties are difficult to obtain. Even when acquired, the accuracy of the data is difficult to guarantee. In terms of a content-centric computing system, a user has the potential to be described by its features. The characteristics of user identities are released from the user’s article. In short, we translate the identity characteristics of users into text features of articles. The structure of the obtained feature vector is shown in Fig. 2.
3.2.1 TF-IDF

In this paper, we use term frequency-inverse document frequency (TF-IDF) [42, 43] to generate the text features of comments. TF-IDF is a common weighting technique in information retrieval and data mining. If some words appear in a text many times, these words will appear frequently in the same type of text. Therefore, if the feature space coordinate system takes the term frequency (TF) word frequency as a measurement, it can reflect the characteristics of the same type of texts. TF-IDF also considers the difference between different types of words. The algorithm believes that if the word frequency in different texts is small, its ability to distinguish different categories is great. For this reason, the inverse text frequency (IDF) concept is added to the algorithm. The product of TF and IDF is considered as the value measure of words in a feature space coordinate system.

Next, we introduce how to compute TF and IDF. TF is calculated as 
\[ tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \]
In this equation, \( n_{i,j} \) represents the number of occurrences of word \( i \) in a text \( j \). \( \sum_k n_{k,j} \) represents the number of occurrences of all words in a text \( j \). Eq. 1 shows how to compute the value of IDF. In Eq. 1, \( |N| \) is the total number of texts in the database. \( \{|j: t_i \in d_j\| \} \) is the number of texts that contain word \( i \). Note that the value of \( \{|j: t_i \in d_j\| \} \) may be zero if the word \( i \) does not appear in the database. Therefore, we usually use \( 1 + \{|j: t_i \in d_j\| \} \) to replace this part.

\[
idf_i = \log \frac{|N|}{|\{j: t_i \in d_j\}|} \approx \log \frac{|N|}{1 + |\{|j: t_i \in d_j\|\}|} \tag{1}
\]

After the acquisition of TF and IDF, we use Eq. 2 to calculate the weight of word \( i \). A high frequency of word \( i \) in a particular text \( j \) and a low frequency of the word in the entire collection of texts can produce a high TF-IDF. As a result, TF-IDF tends to filter out common words and to retain important words. Note that Eq. 1 and Eq. 2 are mentioned in [42].

\[
w_{i,j} = tf_{i,j} \times idf_i = \frac{n_{i,j}}{\sum_k n_{k,j}} \times \log \frac{|N|}{1 + |\{|j: t_i \in d_j\|\}|} \tag{2}
\]

3.2.2 Feature extraction

In social networks, each group has its own glossary; the size of this glossary may be very large. Furthermore, the glossary used to represent the group is created by all users in that group. The glossary can only be updated by a trusted third party after the group is established, and the individual user has no ability to update the glossary. The update operations are only performed by social networks. The TF * IDF feature of the user’s shared information in the group is regularly extracted, and the top-ranked words are selected as the new glossary in order from high to low. After the update operation, the trusted third party sends the new glossary through a secure channel to each user. At the same time, the trusted third party updates all users’ features from the articles shared within a period of time.

We now extract the feature vectors for every user in a group. First, the group’s glossary contains the following words \{cryptography, security, privacy, sharing, agreement, content, computing\}. When a user wants to join this group, the trusted third party calculates the weight of all meaningful words and sorts them from high to low. These words are from information previously shared in social networks. The extraction algorithm is based on the value of TF * IDF. The highest ten words are selected as the user’s identity. After selecting these words, the third party sends the phrase group to the user, and the third party does not receive any other information. After the user receives
his own features, the user compares his identity feature with the glossary of the group. If the user’s identity feature contains the group feature, the corresponding vector is set to 1. If the user’s identity feature does not contain a certain group feature, the corresponding vector is set to 0. This process produces a feature vector of only 0 and 1. For example, the extracted feature of one user is \{traveling, security, privacy, food, university, computer, transportation\}. Then, the user can get his feature vector as \([0, 1, 1, 0, 0, 0, 1]\). In this step, we consider words that have the same root as the same word, e.g., computer and computing are equal in our model. After each user in this group has obtained its feature vector, we use an arbitrary value to represent the user’s feature vector. We also use a matrix to represent the user group. In this matrix, a column represents a user. The form of the matrix is shown as:

\[
\begin{bmatrix}
0 & 1 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 & 1 \\
0 & 1 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 \\
\end{bmatrix}
\]

### 4 GROUP USERS AUTHENTICATION SCHEME

In this Section, the group users authentication scheme in social networks, which is available to attach multiple users to one hobby group in the network, is introduced in detail. Note that the users in each group are controlled by the TGM, which takes the social service provider as a representative. In addition, the number of TGMs is greater than one. Under our assumptions, the identity numbers of users in a group are delivered to the users after user registration is complete. We emphasize that the mutual communication between users in the initialization phase is not used in our scheme, preventing a data collision problem in the network. Users communicate with the TGM according to their identify numbers and the TGM’s identify number and gather all the secret information to generate the proof. The authentication process is completed when all the users in the group are verified successfully, which means that these users are credible so that the network is safe and well arranged. In addition, unusual situations are considered. For example, when some of the users are broken or damaged, the TGM can generate a proof containing other users’ secrets. Consequently, the TGM acquires the information that some specific users need to be repaired, but the group remains safe, which is practical in real systems.

Next, we introduce the definition of this problem. We assume that a set of large groups needs to be verified, where each group is attached to more than one user in its different regions. Note that the users combined within the same group are considered as one user group. The group is arranged to be successfully authenticated after all the users are fully verified, which guarantees that the whole group is considered to be safe. The proposed group user authentication scheme should meet most of the security requirements without affecting its efficiency or increasing the cost. Moreover, the TGM should be provided with the feedback of all the users in the group.

#### 4.1 Design of the scheme

We divide the group authentication scheme into four phases for a better description. In the initialization phase, users are initiated with the TGM’s identify along with the identity numbers as soon as they are attached to the social group. The total number of users in the network is \(N\). The TGM acquires \(N_{TGM}\) users identities and combines them with the random numbers received from the server in the user acquisition phase. The proof of group \(G\) is generated in the group authentication phase, which links the entire group. In the verification phase, the proof is delivered to the server. It is necessary to emphasize that the communication between the server and the TGM is performed via a secure channel, while that between the TGM and the users is not safe and is vulnerable to various attacks. The notation used in our proposed scheme is provided in Table 1.

#### 4.1.1 Initial Phase

User initialization is conducted in this phase. As we describe above, one group is assumed to be combined with several users to ensure that the group is credible. The group information, as well as the pseudo-identify \(ID_{user_i} = H(ID_i || f_{user_i})\), is input in the memory of the users. The pseudo-random numbers \(r_{ID}\) of the users are stored in the server database. Note that the user referred to here can be a reusable user, which means that the user can be attached to another group as necessary.

#### 4.1.2 User Acquisition Phase

In this phase, the users that respond to the TGM are registered after verification. According to our design, it is
not necessary for the social group to collect all $n$ users. The network allows authentication towards more than one user. We assume there are $N_{G_j}$ users in one group that are successfully registered. In extreme situations, only one user that passes the authentication of the network can prove that the group is not invalid. However, whether the group is credible remains unknown.

The detailed steps of this phase are as follows:

- The server generates a pseudo-random number $r_{TGM_i}$ for $TGM_i$. Then, the server chooses the group to be authenticated and delivers $(ID_{TGM_i}, GID_{Gi}, r_{TGM_i})$ to $TGM_i$ through a secure channel.
- $TGM_i$ computes $M_{G_i} = (GID_{Gi} \oplus ID_{TGM_i}) + (S_{G_i} \lor r_{TGM_i})$ and sends $(ID_{TGM_i}, M_{G_i})$ to all the users in the group.
- The user $user_i$ of group $G_i$ checks the validity of the received $M_{G_i}$ and obtains its own pseudo-random number $r_{user_i}$ from the TCS. Then, $user_i$ computes $N_{user_i} = [M_{G_i} - (GID_{Gi} \oplus ID_{TGM_i})] \lor r_{user_i}$, $Q_{user_i} = ID_{user_i} \oplus (S_{G_i} \lor r_{user_i})$ and $R_{user_i} = r_{user_i} \ast p$. Next, the user $user_i$ responds $(N_{user_i}, Q_{user_i}, R_{user_i})$ to $TGM_i$.
- After receiving the information from the user, the $TGM_i$ checks the validity of the acquired $N_{user_i}$, and computes $ID_{user_i}$ from $Q_{user_i}$. When all the users have been collected or the timer has expired, the $TGM_i$ sends $(ID_{TGM_i}, GID_{Gi}, ID_{user_1}, \ldots, ID_{user_i})$ to the server in a secure way.
- The server compares the received information and the information stored in memory. If this comparison succeeds, the server replies to the $TGM_i$ with $(ID_{TGM_i}, GID_{Gi}, R_{user_1}, \ldots, R_{user_i})$, where $R_{user_1}, \ldots, R_{user_i}$ denote the pseudo-random numbers calculated for each user separately. The $TGM_i$ stores all this information in its database for the next authentication phase.

The characterization of the user acquisition phase is described briefly in Fig. 4.

### 4.1.3 Group Authentication Phase

This phase is the most important part of the authentication phase and focuses on gathering the secrets of each user and combining this information according to the user’s sequence number $ID_{user_i}$. Instead of using the user’s real sequence number $ID_{user_i}$, we compute a temporary identifier $TempID_{user_i}$, which is the product of a random number $r_{user_i}$ and $ID_{user_i}$. The TempID$_{user_i}$ of user $user_i$ is exchanged between the user and the $TGM_i$ during the communication, which shows sufficient resistance to the tracing problems of smart users. In addition, the random number $r_{user_i}$ is randomly selected by the user and is upgraded after every authentication progress. In the group authentication phase, the $TGM_i$ does not terminate the process when the requested user fails to respond. In other words, the $TGM_i$ skips to the next normal user after the timer expires. This approach prevents intervention in the whole system by abnormal users or user groups, which makes the authentication process more efficient.

The main steps of the group authentication phase are as follows:

- $TGM_i$ computes $r'_{user_1} = \hat{e}(R_{user_1}, H1(ID_{user_1}))$ and $TempID_{user_1} = (GID_{Gi} \oplus ID_{user_1}) \lor (r'_{user_1} \oplus ID_{TGM_i})$. Then, the $TGM_i$ delivers $(Q_{user_1}, TempID_{user_1}, Onec, R_{user_1})$ to the first user in group $G_i$.
- The first user $user_1$ in $G_i$ checks the validity of $TempID_{user_1}$ received from $TGM_i$ and computes $Prof_{GID_{Gi}} = TempID_{user_1} \lor (r'_{user_1} \oplus$...
Fig. 4: Group authentication phase

Fig. 5: Verification phase

5 SECURITY ANALYSIS

In this Section, we analyze the security properties of our scheme to prove that our scheme is reliable. The security of our scheme is based on the ECDLP [44, 45].

5.1 Security against passive attacks

In the following lemma, we prove that the matrices $A_i$, $i = 1, 2, ..., v$, generated from the user’s comments, have a uniform distribution on $\mathbb{Z}_q^{t \times r}$, which is impossible to compute from another matrix.

Theorem 1. Let $A_k$ and $A_i$ be the chosen matrix, which is generated in Section 3, and $x \in \{0, 1\}^*$ be a random vector. In this lemma, $A_kx$ cannot be calculated from $A_ix$ in polynomial time, where $1 \leq k \neq i \leq v$.

Proof. Algorithm $A$ takes $A_ix$ as the function’s input and outputs $A_kx$, where $A_k$ and $A_k$ are calculated in Section 3. $x$ is a random vector from $\mathbb{Z}_q^{t \times r}$ and is randomly chosen.
from \(\{0,1\}^r\). In addition, \(A\) can obtain the result with a non-negligible probability. By applying algorithm \(A\), we propose an algorithm \(B\) that takes \(Mx\) and \(M \in \mathbb{R}^{r \times r}\) as input. Firstly, \(B\) selects a random matrix \(A_k\) from \(\mathbb{Z}_q^{r \times r}\). According to lemma 2 in [46], \(A_k\) has full row rank with overwhelming probability. Next, \(B\) calculates a matrix \(E \in \mathbb{Z}_q^{r \times (r-t)}\). \(E\)'s columns form a basis for the null space of matrix \(A_k\) (i.e., \(A_k E = 0_{r \times (r-t)}\)). Then, \(B\) computes \(A_i = M [A_k^* E]^{-1}\). Note that \(A_i\) is uniformly distributed on \(\mathbb{Z}_q^{r \times t}\). \(A_k^* \in \mathbb{Z}_q^{r \times r}\) is the pseudo-inverse matrix of \(A_k\), which can be calculated as \(A_k^* = (A_k A_k^T)^{-1}\) because \(A_k\) has full row rank. The matrix \([A_k^* E]\) is a reversible matrix because \(A_k^*\) and \(E\) have full column rank. No column in matrix \(A_k^*\) is a multiple of a column in \(E\). Furthermore, \(A_k A_k^* = I_t\), where \(I_t\) has no zero columns. Let \(y = [A_k^* E]x\). On input \(Mx = Ay\), the algorithm \(A\) outputs \(k = A_k y = A_k [A_k^* E]x = [I_t A_k E]x = x_1 + A_k E x_2 = x_1\), where \(x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}\). Therefore, on input \(Mx\), the algorithm \(B\) can only get the first part of \(x\), which cannot be used to obtain all the information about the matrix.

5.2 Security against active attacks
In addition, the scheme is resistant to various attacks [47]. Social networks have great potential in daily communication. Because they are convenient for data sharing. Note that the computational resources consumed by the users in the group authentication scheme are very small, which is consistent with the requirements of low cost and low power qualities of mobile users. Furthermore, each communication is added to the protection of the transfer of information to meet the strong privacy protection of users. Moreover, the scheme can be used to track attackers. The TGM controls and monitors the entire network changes in the group authentication scheme, thereby improving security and practicality in social networks.

5.2.1 Resistance to replay attack
In our scheme, replay attack [18] is prevented by using pseudo-identities. As we describe above, in the initial phase, the feature vector of \(user_i\) is generated by the TGM, and \(user_i\) randomly selects \(r_{user}i\) to compute \(R_{user}\). In our model, the users take full control of the group authentication scheme. As for data sharing among users in a social group, the pseudo-identity is calculated by the TGM and itself from the phrase group. The content of each share is not the same, so the pseudo-identities at different times are not equal. In addition, the user itself generates the random number \(r_{user}i\) in the initial phase in every session. The reuse of the previous messages cannot pass the information of both the TGM and the other users, so the replay attack is prevented.

5.2.2 Resistance to eavesdropping
The attackers may attempt to obtain shared data through eavesdropping [45] during the communication process, which is a common way for attackers to damage social networks. In our scheme, the ID of the users, as well as the secret keys, are under encryption. We apply \(ID_i || f_{\text{user}}i\) to the communication between users so that it is not easy for the attackers to acquire the privacy information of a particular user through eavesdropping. Moreover, the TGM controls the generation of all pseudo-identities, and the data exchanges between the TGM and normal users are assumed to be safe. As a result, the entire social networks in our scheme are resistant to eavesdropping.

5.2.3 Resistance to physical attack
We assume that the attackers are able to destroy normal users and to obtain all the secret information of the destroyed user. The attackers may even manage to clone one user and use it to cheat the TGM. However, in our scheme, one group is attached to a group of users. The group can be authenticated with several malicious users in the network. To interrupt the authentication process, the attacker has to damage all the users in the social group, which is difficult and worthless. As the important facility of the entire social network, the TGM is assumed to be well protected by outsourcing to a credible third party. As a result, physical attacks towards the TGM are not considered in this paper. We attach significance to the security requirements of the users rather than those of the TGM.

5.2.4 Resistance to tracing attack
Tracing attacks [48] towards users, especially users who have low security levels, remain a big threat to both the group and the network. We classify tracing attacks into two types: tracing during the session and tracing between two successful sessions. On the one hand, to prevent tracing attack during one session, the messages delivered between the users and the TGM are all under encryption. The use of a pseudo-identity provides strong protection through tracing during the session. In addition, the \(ID_{user}\), of \(user_i\) is changed along with data sharing. On the other hand, random number \(r_{user}\) is used to change the \(R_{user}\), of \(user_i\) after every successful authentication session. As a result, tracing between two successful sessions is impossible.

5.2.5 Resistance to de-synchronization attack
In a de-synchronization attack [38], an adversary disturbs the interactions between a user and the TGM by intercepting or blocking the packages. In our scheme, the real ID of the node is not exposed in the authentication process. Instead, pseudo-identities are used. As a result, it is not necessary for users to update their secrets. In addition, the users are relatively independent from each other. The compromising of one user will not affect the entire group and the network. In this situation, if one or more users are blocked several times, the users can also contact the TGM, i.e., the users in the social group will not suffer from de-synchronization attacks according to our design.

5.2.6 Assurance forward security
In our scheme, if the adversary obtains the secret information of the sensor node in the present round, the secure information of the previous rounds will not be leaked to the adversary [31]. The \(ID_{user}\) in every round is generated and valid only in the present round, which means that the secret information is not related to the previous round. Therefore, forward security is provided in the proposed scheme.
TABLE 2: Security comparison among our scheme and existing schemes

<table>
<thead>
<tr>
<th></th>
<th>Jiang et al.’s scheme [18]</th>
<th>Shen et al.’s scheme [33]</th>
<th>Horng et al.’s scheme [34]</th>
<th>Our scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replay Attack</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Eavesdropping</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Physical Attack</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Tracing attack</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>De-Synchronization</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Forward Security</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

5.3 Security comparison

In this subsection, we compare security properties of the proposed scheme with [18], [33] and [34]. The comparison result is given in Table 2. In [18], the old password is shared between the gateway node and the user, thus the forward security cannot be guaranteed. In [33], the secret information is stored in local nodes, therefore the physical attack and the tracing attack as well as the de-synchronization attack cannot be resisted. In [34], signatures can be forged by adversaries, so the eavesdropping cannot be withheld. From Table 2, it is easy to find that our scheme can satisfy all security properties.

6 Performance evaluation

In this section, we compare the proposed scheme with Horng’s scheme [34] in the view of their computational cost. Note that all the point multiplication and exponentiation for the authentication cost are operated at the user side in [18] and [33], while the authentication cost are only operated at the server side in Horng’s scheme and our scheme. Hence, in the performance evaluation, we only compare our scheme with Horng’s scheme to show our advantages.

According to the design of the group user authentication scheme, each user should receive 2 messages from the TGM. After receiving the 2 messages, the user performs necessary operations including point multiply on the elliptical curve and modified Weil pairing computation. The computational cost in one group user authentication scheme is mainly a result of pairing computation, point multiplications and exponentiation operations. The communication cost mainly comes from the information exchange between users in the group user authentication phase. For the convenience of evaluating the computational cost, we let $T_{H1}, T_{H2}, T_{XOR}, T_{AO}, T_{CO}, T_{PC}, T_{PM}, T_{EO}$ be the time of implementing a hash to point function, a hash function, a XOR operation, an addition operation, a concatenation operation, a modified Weil pairing, an elliptic curve point multiplication and an exponential operation. To describe the computational cost of different schemes, we refer to the simple method in [49]. For example, in the user acquisition phase of the group authentication scheme, $user_i$ will initially calculate $N_{user_i}, Q_{user_i}$ and $R_{user_i}$. The computational cost of modified Weil pairing and elliptic curve point multiplication for each user is $T_{PC}$ and $T_{PM}$, respectively. In the following processes, $user_i$ will receive 2 messages from the TGM. We can easily see that the number of received messages by $user_i$ is 2. In the group authentication phase, $user_i$ will compute $(TempId_{user_i}, C_{user_i}, Prof_{GIDi})$. The total computation cost of each user in the user acquisition phase is then $2T_{XOR} + 3T_{AO} + T_{PM}$. In the group authentication phase, the method of calculating the computation cost of the three operations is the same as in the first round. We can obtain the total computation cost as $1T_{H1} + 1T_{PC} + 2T_{XOR} + 2T_{AO}$. Finally, the computational cost in group authentication scheme is $T_{H1} + T_{PC} + T_{PM} + 4T_{XOR} + 5T_{AO}$ for each user. The comparisons between our scheme and Horng’s scheme are shown in Table 3.

The PBC Library project was funded by Stanford University. This project was designed to develop a tool for researchers to verify the accuracy of schemes by targeting the design of the schemes. To verify the performance of our scheme, we provide an experimental evaluation of the proposed scheme. Our experiments are simulated using C programming language with the PBC library and the GNU multiple precision arithmetic (GMP) library on a Ubuntu operating system with Intel Core Xeon E5-2650M processors running at 2.60 GHz and 8 G memory, Ubuntu 14.04 X64. From our simulation, the time required to perform the hash function, hash to point function operation, a pairing computation, point multiplication, exponentiation and addition is approximately 8.065 ms, 0.005 ms, 2.067 ms, 3.069 ms, 3.072 ms and 0.016 ms. The total time cost of each user in our scheme is computed by $8.065 + 2.067 + 3.069 + 5 \times 0.016 = 13.881\text{ms}$. However, the time cost of each user in Horng’s scheme is $38.699\text{ms}$.

We first simulate each phase of the group user authentication scheme. The computation cost of the TGM does not require resources from the group users. Therefore, we do not take the computation cost of the TGM into consideration. The comparison of different phases in our scheme is shown in Fig. 6. The initial phase is the preparation stage, and most operations are performed by the TGM. Fig. 6 shows the time consumed in two phases of our group user authentication scheme.

In addition, Fig. 6 is used to describe the computational cost of different phases of our scheme including the user acquisition phase and the group authentication phase. We simulated the computational cost of each phase from the first simulation to the 50th simulation. Note that simulation results in different simulation times have some deviations. Hence, in order to better present simulation results in Fig. 7, we use the average value based on 50 simulation times.
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TETC.2017.2779163, IEEE Transactions on Emerging Topics in Computing

7 Conclusion

In this paper, we propose a content-centric group user authentication scheme to achieve data sharing and to protect users’ privacy by solving the problem of generating pseudo-identities based on the word’s TF * IDF value. The proposed scheme is based on the design of a group manager and the pseudo-identity of users in one social group. In our scheme, the content-centric computing system is used to calculate the user’s feature vector, which can be used as a part of user’s pseudo-identity, so as to protect the user’s real ID during the group authentication. According to the security analysis, our scheme can resist multiple attacks. When the attacker wants to infiltrate the network, the TGM can detect it with the help of normal users. Meanwhile, the group user authentication scheme is not interrupted by the attacker. In addition, the computational costs of our scheme are lower than those of the existing schemes. In the further work, we plan to improve the scheme by solving the following problems.

- Deleting the information of users who want to leave the group.
- Keeping in sync between the TGM and the TCS.
- Determining the trusted group manager from more than one group managers.

## References


Jian Shen received a B.E. degree from Nanjing University of Information Science and Technology, Nanjing, China in 2007 and M.E. and Ph.D. degrees in Computer Science from Chosun University, Gwangju, South Korea in 2009 and 2012, respectively. Since late 2012, he has been a professor at Nanjing University of Information Science and Technology, Nanjing, China. His research interests include computer networking, security systems, mobile computing and networking, ad hoc networks and systems, and ubiquitous sensor networks.

Anxi Wang received a B.E. degree in 2016 and is currently working toward an M.E. degree at Nanjing University of Information Science and Technology, Nanjing, China. He focuses on key agreement protocols in networks and security systems. His research interests include ad hoc networks and systems, network security, information security, and wireless sensor networks.

Chen Wang received a B.E. degree in 2016 and is currently working toward an M.E. degree at Nanjing University of Information Science and Technology, Nanjing, China. He focuses on directional antennas in ad hoc networks. His research interests include ad hoc networks and systems and wireless sensor networks.

Jiguo Li received a Ph.D. degree from Harbin Institute of Technology in 2003. Since 2003, he has been with Hohai University. He is currently a Professor with the College of Computer and Information Engineering. His major research interests include information security and cryptography, network security, wireless security, and trusted computing. He has published more than 70 academic papers and two books. He has served as a PC Member of several international conferences and as a Reviewer of international journals and conferences.

Jian Shen

Chen Wang

Yan Zhang

Jiguo Li

Anxi Wang

10.1109/TIFS.2017.2705620.