IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 31, 2017, accepted October 1, 2017, date of publication November 8, 2017,
date of current version November 28, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2767285

Authentication With Block-Chain Algorithm and Text Encryption Protocol in Calculation of Social Network

RUIGUO YU[1,2], JIANRONG WANG[1,3], TIANYI XU[1,3], JIE GAO[1,3], YONGLI AN[1,3], GONG ZHANG[1,3], AND MEI YU[1,3]

[1]School of Computer Science and Technology, Tianjin University, Tianjin 300350, China
[2]Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin 300350, China
[3]Tianjin Key Laboratory of Advanced Networking, Tianjin 300350, China

Corresponding author: Mei Yu (yumei@tju.edu.cn)

ABSTRACT Community detection is an important aspect of social network analysis, but social factors such as user intimacy, influence, and user interaction behavior are often overlooked as important factors. Most of the existing methods are single classification algorithms; multi-classification algorithms that can discover overlapping communities are still incomplete. In former works, we calculated intimacy based on the relationship between users, and divided them into their social communities based on intimacy. However, a malicious user can obtain the other user relationships, thus to infer other users interests, and even pretend to be the another user to cheat others. Therefore the information users concerned about needs to be transferred in the manner of privacy protection. In this paper, we propose an efficient privacy preserving algorithm to preserve the privacy of information in social networks. First, during expansion of communities on the base of mining seed, in order to prevent others from malicious users, we verify their identities after they send a request. We make use of the recognition and nontampering of the block chain to store the user's public key and bind to the block address, which is used for authentication. At the same time, in order to prevent the honest but curious users from illegal access to other users' information, we do not send plaintext directly after the authentication, but hash the attributes by mixed hash encryption to make sure that users can only calculate the matching degree rather than know specific information of other users. Analysis shows that our protocol would serve well against different types of attacks.

INDEX TERMS Information protection, block chain, hash encryption, text encryption protocol.

I. INTRODUCTION

Although social network analysis has now reached a relatively mature stage, a large number of scholars continue to improve the convenience and security of social networks. It is observed that in complex social networks, there are multiple overlapping communities [1]. Numbers of community detection methods exist, one of which is to detect communities by analyzing user properties and interests [2]. However, as social networks have become more and more complex, we can no longer detect communities according to these simple rules. We need to reconsider more factors in terms of circumstances. To deal with that, in user centric social networks, we detect social communities according to user influence, user relation and interaction, also we practice personalize recommending based on semantic analysis and statistical analysis. Experiments show that our method could lead to a better detection effect in finding communities and a better recommending effect in former work [3].

Social network applications reflect real world to cyberspace, thus lead to privacy leaks while detecting social circles according to user information. User private information is necessary when social circles are to be detected. But once this privacy information is illegally obtained, criminals will be able to obtain the relationship between users, infer other user interests, and use other user personal information. The consequences will be unbearable to contemplate.

To avoid that risk, we need to validate the identity of user in the process of building social circles by relationship of users. In this process, public key cryptography is usually used. But this method would fail on one condition: when someone forges the identity and key to match, the user cannot be identified effectively, resulting in his private information being leaked. Digital certificates can solve this problem [4], but publishing digital certificates require specific agencies, which make them inconvenient to use. Especially in the social network, since the amount of users is huge, it is impractical to make a digital certificate for each person. Therefore, we propose the use of block chaining technology to protect the identity of users and the security of their corresponding keys, so as to resist the attacks from semi-honest users and malicious users. In this paper, we improve CMCR [3] algorithm through two aspects i.e. user authentication and text encryption, so that it can better ensure the security of user information. In terms of user authentication, we propose the Authentication with Block-chain algorithm to verify the identity of user while not causing private information leakage. In terms of recommendation text, we propose a protocol combining RSA algorithm to prevent users illegally acquiring information.

The paper is organized as follows. In the second part, we introduce the related work of the thesises and their limitations. In the third part, we introduce some related technical knowledges and related definitions of defense attack in this paper. In the fourth part, we propose Authentication with Block-chain algorithm to introduce social circle development and to expand privacy protection in the social circle and recommendation process. In the fifth part, we evaluate our model and test security. The last part is summary and outlook of the work.

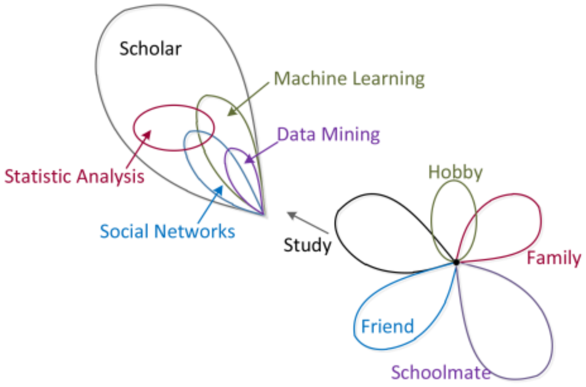


FIGURE 1. **The overlapping communities.**

II. RELATED WORK

In dealing with social communities detection, most of early approaches [5]–[7] are based on the entire network structure, so they don't function well on detecting overlapping communities of a certain user. And most of these algorithms are classification algorithms. Fortunato *et al.* [8] summarize the methods of community detection, and illustrate the features of overlapping communities as in Figure 1. Although he focus on the overall structure of the network, it does not provide a way to detect the community of a particular user. We propose an algorithm named CMCR for a user centric network in Paper [3], a novel algorithm based on Clique Theory [9], PageRank, LDA, and TF-IDF used to solve the problem. Experimental results show that the proposed algorithm can outperform baseline algorithms in some common criteria. However, when designing this algorithm, there is a slight lack of user privacy and user information protection.

There are two main privacy protection mechanisms in user profile based user matching. A mechanism treats user profiles as collections of attributes and then matches them according to the set of attributes [10], [11]. Another mechanism allows user profiles to be matched by vector dot products by regarding as user profiles as vectors [12]. The two mechanisms mentioned above are based on public key cryptography and homomorphic encryption technology, so the computation cost is very high. And these two mechanisms require a trusted third party organization [13], which is hard to implement in social networks. We do not use pre-configured trusted third party organizations to encrypt using attribute information, which is not a direct match between users.

Block chain technology and distributed Sub Ledger have attracted much attention and led to many projects in different industries [14]. Paper [15] uses block chaining technology to run the cash system in a peer to peer environment and prevents double payment, and solves the problem that a trusted third party handle the electronic payment information. Paper [16] applies block chain technology to social networks, uses reliability scoring to improve the system, and analyzes attacks rather than using PoW(prove of work) to protect them. But in the process of encrypting configuration file, RSA is the first public key cryptosystem to prevent applications, and its security has always been the focus of cryptography research. It is widely used in various applications in the security field. RSA and other related technologies, combined with biological development of new technologies have become a new research point [17].

III. PRELIMINARIES

A. BLOCK CHAIN

Block chain is a kind of technique realization of the electronic currency book system by peer-to-peer, it can record every bitcoin transaction records without a center server in a network system, and it is maintained by participants. No one can change the contents of the block chain without authorization, thus it has very stable security for its holder. It allows any two users to trade directly without a trusted third party mechanism. The block chain records all transactions that occur in the bitcoin system, and once the transaction information is recorded, it is permanently stored and cannot be changed. In a bitcoin system, only if 51% or more of the nodes are controlled by attackers, attackers could launch attacks on the system. Since most nodes are controlled by honest network nodes, attacks are very difficult to implement, thus the block information in the block chain is trustworthy. The anonymity of participants in the bitcoin system ensures the security of their privacy. Participants can either voluntarily leave or re-enter the bitcoin system by receiving the longest workload proof chain to obtain transaction information that occurs when leaving the system.

B. ATTACK MODEL

1) SEMI-HONEST MODEL

In this model, semi honest members are also called passive attackers. In the process of multi-party computation, a semi-honest member fully abides by the implementation of the agreement, neither withdraws from the agreement, nor tampers with the results of the protocol. He or she may retain some intermediate results in the implementation of the agreement and attempt to analyze and derive input data from other members through these intermediate results.

2) MALICIOUS MODEL

In the model, malicious attackers are also active attackers. In the calculation process, a malicious attacker cannot follow the protocol process execution, interrupt protocol operation process, and collude with the intermediate results or modify the agreement with other parties.

IV. PERFECTION OF SOCIAL CIRCLE DETECTION

A. SOCIAL CIRCLE DETECTION

Overlapping community detection would be done by mining seeds and expansion. Social circle expansion would be conducted by extending the seed set $Seed(c)$ through two algorithms of Closeness Seed Expansion (CSE) and Influence Social Community Expansion (ISCE) algorithm.

According to the user information and the user relationship, seed mining can obtain the seed set $Seed(c)$ by an algorithm named K-clique-community Seed Mining (KSM) algorithm to solve that problem. KSM takes the two important theories of Clique Theory [8] into account to implement the mining. In the case of all cliques detected, KSM implements the mining according to Clique theory, constructs the group overlap matrix M, and calculates the adjacent matrix M' of the undirected graph. Then, the depth of the connected subgraph in M' is searched first, and the last maximum clique $Seed(c)$ is obtained.

The first algorithm used in social circle expansion is the CSE algorithm, which uses a greedy strategy to prioritize the closeness of the central user. $Seed(c)$, as the initial community of CSE algorithms, is made greedy expansion based on the affinity feature, and added closer nodes to the community. Social circles with size not bigger than k would be removed, where k is a parameter in K-clique-community algorithm.

The second algorithm used in social circle expansion is ISCE algorithm, which focuses on the user's impact on the network. First, we calculate the user's impact, according to which the candidates are sorted in the second-level relationship in the ISCE. The Calculation process of user impact is shown in the Figure 2. Then the classic modular function Q is adopted as the standard. If the present social structure would form a better social structure with a new added node, the modularity will increase. Then that node would suit the community for sure. By analysis, we believe that the number of users replying to a message can indicate his interest in message. When computing user influence, we use the idea of similarity with random walk in PageRank to represent user relations. The similarity $Sim(i, j)$ between users is calculated by the cosine similarity of the eigenvector of the message interest. The message interest feature vector is defined in Definition 1, where id represents the message's ID, and $cn(u, id)$ represents the answer number of the user u reply message ID.

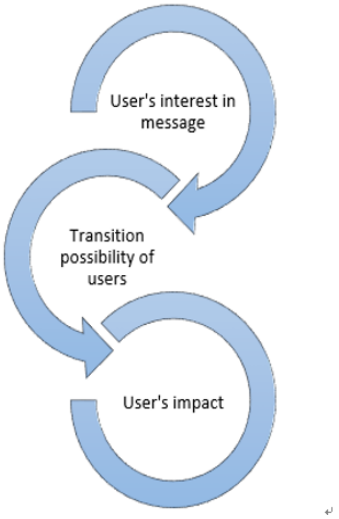


FIGURE 2. Calculation process of user impact.

Definition 1. Message Interest Feature Vector: define $V(u)$ as the message interest eigenvector, as shown in Equation (1):

$$V(u) = [id_1 = cn(u, id_1), id_2 = cn(u, id_2), \cdots, id_t = cn(u, id_t)], t = |M(u))|, id_i \in M(u) \tag{1}$$

The transition possibility of users stands for the ratio of information which User i is interested in from the whole information he gains. The method of calculation is shown in Equation (2). The molecule is the amount of information received by the user i from the user j. The denominator is the total amount of information received by the user i.

$$P_{i,j} = \frac{M(j) \times Sim(i,j)}{\sum_{n \in V_1} |M(n)||Sim(i,n)|}, i, j \in V_2. \tag{2}$$

Finally, we calculate the user's impact by Formula (3), where q is set by experiments in PageRank.

$$PR(i) = \frac{1-q}{V_2} + \sum_j PR(j) \times P_{i,j}, i, j \in V_2 \tag{3}$$

B. AUTHENTICATION AND RELATIONSHIP ENCRYPTION

CSE algorithm is based on the intimacy of users. In sociology, if two people have more common friends, the relationship between the two users is more intimate. So we calculate intimacy by calculating two users who follow each other. However, in social network implementation environment, to conduct social circle detection, a user would have to know other users' private information, in order to prevent other people illegally obtain user information, any two party users should be verified their identity before the process of communication. Therefore, before a user acquires information of the other user relationship, we need to authenticate identity of that user, in case a malicious user gets the relationship of the users illegally, and infer the user interests and preferences.

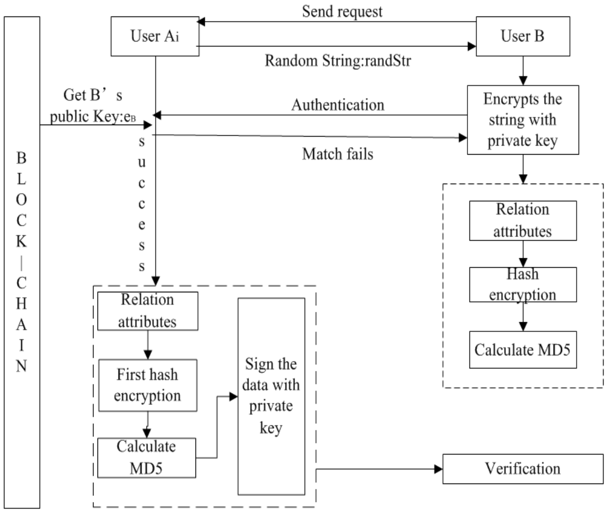


FIGURE 3. **Authentication model flow.**

In this section, we use the Authentication with Block-chain algorithm to authenticate users. We create a pair of private key and public key for each user. To ensure the security of the key, we store the public key chain according to the block chain constructed by the bitcoin system. The model flow is shown in Figure 3.

In generation of public key and private key, we use the state-of-art RSA algorithm.

Definition 2: suppose $N = p^r q$, where r is a positive integer, p and q are prime numbers random generated, and the corresponding plaintext space P, cipher-text space C and key space $\mathbb{Z}_n$ can be defined as $P = C = \mathbb{Z}_n$ respectively to satisfy the Equation (4) (5).

$$K := (N, e, d) | ed \equiv 1 \ mod \ \Phi(n) \tag{4}$$

$$\Phi(n) = p^r(p-1)(q-1) \tag{5}$$

By Definition 2, we get private Key e and public Key d respectively. The encryption $e_k(m)$ of plaintext $m \in P \ \mathbb{Z}_n \rightarrow \mathbb{Z}_n$ as shown in Formula (6).

$$e_k(m) := m^e \ mod \ n \tag{6}$$

We implement authentication between social circle and user with the Authentication with Block-chain algorithm. Assume that A_i belongs to the user set $U(A_1, A_2, \cdots, A_n)$ and B is the candidate user node to be added into that user set. We calculate the intimacy between user B and each user in social circle. The algorithm is as Table 1.

TABLE 1. **Authentication with block-chain algorithm.**

```
Algorithm 1:Authentication with Block-chain
Input:
    $A_i$'s public key $e_{A_i}$ and private key $d_{A_i}$, B's public key $e_B$ and private key $d_B$. $e_{A_i}$ and $e_B$ are stored on block-chain BC;
Output:
    verification result of B, 1 represents success, 0 represents failure

1. User B sends requests to user $A_i$;
2. $A_i$ generates a validated random string randStr;
3. $A_i$ sends randStr to user B;
4. B encrypts the string with his private key and sends it to $A_i$
5. User $A_i$ obtains the B's public key $e_B$ from the block chain BC and
   decrypts the string to get the $m'$.
6. If $m! = m'$
7.     Return 0; //match failure
8. $A_i$ encrypts the attributes by hash, then calculates the generated hash
   by MD5 as message digest $Md_{A_i}$.
9. $A_i$ encrypts his/her own digest with the private key $d_{A_i}$ and sends the
   $E(Md)_{privKey}$ to B
10. B decrypts through the public key $e_{A_i}$ of A
11. If decryption successful
12.     B matches with $A_i$ by his own digest
13.     Return 1//match successfully
```

First, users create their own public and private key pairs, and use their own block chain system to store the public key on the block chain nodes. A block chain consists of two nodes forming a private block chain. A certain user keeps his private key, and his public key is stored in the established block chain binding with address. System gives users permission of the smart contract in block chain to ensure that users can access the public key of the other users through address.

Second, user A_i needs to encrypt his relationship attributes before sending his information to B. He needs to prepare a relational attribute set, his own private key, his block chain address, and the public key. Suppose that the attribute set is $S^{A_i} = (s_1, s_2, s_3, \cdots, s_n)$, we first compute the hash value $H(S^{A_i}) = (h_{s_1}, h_{s_2}, h_{s_3}, \cdots, h_{s_n})$ of the attributes S^{A_i}. Because the generated hash string is 64 bit, and the length is slight longer, the time overhead required for matching is relatively large, and the security of one time hash encryption is slightly worse at the same time. So we encrypt it twice on basis of $H(S^{A_i})$ and get MD5 digest: $Md^{A_i} = (md(h_{s_1}), md(h_{s_2}), md(h_{s_3}), \cdots, md(h_{s_n}))$. Then, A_i encrypts its own MD5 digest with its own private key and sends the result $E(Md)_{privKey}$ to B.

Finally, B hashes his relational attributes before it is matched, and produces its own MD5 digest. After receiving the information sent by A_i, B decrypts the signature of A_i by decrypting it with A_i's public key. When the signature is true, B matches the A_i relationship digest with his own relationship digest. After the success of the match, user B and user A_i have higher intimacy. If the match fails, because A_i does not send specific information about relationship, B does not get the relationship of A_i and the privacy of user A_i can be protected. We calculate similarity $Sim(A_i, B)$ between A_i and B according to digests. They can be regarded as two vectors and the similarity between A_i and B can be calculated as following Formula (7).

$$Sim(A_i, B) = \frac{Md^{A_i} \cdot Md^B}{|Md^{A_i}| \times |Md^B|} \tag{7}$$

If the similarity is high enough, B and A_i would have high intimacy. Otherwise, the closeness of B and A_i does not meet the requirements. Because B receives only a hash value of the properties of A_i, the digest is encrypted twice, and the hash cipher is not invertible, thus user B doesn't know the text of the properties of user A_i.

C. TEXT ENCRYPTION BASED ON CONTENT RECOMMENDATION

To facilitate content recommendation, the server generates a pair of public key e_s and private key d_s, which contributes to users to send their own message to server protectively. The same as the user's public key, the public key of the server and other relevant information of the public key, are stored in the block chain. Thus, when the user encrypts the information, the correct public key can be obtained directly from the block chain, thereby avoiding the threat that a malicious user sends the false public key and decrypts the information.

This process requires two steps: calculation and encryption of recommend content. Encryption algorithm is used twice. The first time is that before calculation of a certain user interest when the server needs to acquire some information about his own profile and posted message of other user. This kind of information needs to be encrypted. The second time is the encryption of the recommended messages.

In the first encryption process, when message M is sent to the system, the user obtains the corresponding system's public key e_s from the block chain. Then, user needs to encrypt M to obtain ciphertext and send the ciphertext to the system. After the system receives the ciphertext, the system uses its own private key d_s to decrypt it.

In the original MCRA algorithm, semantic analysis and statistical analysis were taken into consideration to recommend messages to the central user. This process requires to calculate $P(M|u)$ which is the semantic interest of user u for message M, and $K(M|u)$ which is the information interest statistic of user u for message M. The message is represented by the bag of words model. $P(M|u)$ should be the possibility of each word. We set the maximum value of $P(w|u)$ in all the words in the message to represent $P(M|u)$ of the whole message in MCRA to avoid the problem of decreasing product value as the message length increases. The possibility of the word "w_i" issued by the user $P(w_i|u)$ is defined as Equation (8), where T represents the topic, and T is the collection of topics U.

$$P(w_i|u) = \sum_{t \in T} P(t|u)P(w_i|t) \tag{8}$$

In order to compute $P(w_i|u)$, we use Latent Dirichlet Allocation (LDA) to train the model to achieve the user topic possibility distribution $P(T|u)$ and the subject word possibility distribution $P(V|T)$. The user topic possibility distribution $P(T|u)$ is a vector whose element $P(t_i|u)$ is the probability that target user u released informations are related to each topic t_i of T.

In calculating $K(M|u)$, we first need to compute the user word weight vector $K(u)$, whose constituent elements are the weight $w(v_i, u)$ of the message $v_n x$ in the message set of the target user u. The policy for calculating $K(M|u)$ is similar to that of $P(M|u)$, such as Formula (9). If the candidate message contains no text in the content file of the target user, the similarity will be set to the minimum weight in $K(u)$.

$$K(M|u) = \begin{cases} min\{w(w_i)\}, & v_i \in Vc, \exists w_i \notin V_{targetusr} \\ max\{K(w_0|u), \cdots, K(w_n|u)\}, & \\ & \exists w_i \in V_{targetusr} \end{cases} \tag{9}$$

After calculating the semantic interest $P(M|u)$ and the information interest statistics $K(M|u)$, the two parameters need to be considered comprehensively. We use the idea of weighted mean to introduce the concept of information score $Score(M, u)$, and the information score is shown in Formula (10). After scoring, select the highest score messages M_{max} and send it to the user.

When Messages M_{max} are sent to the user, the system obtains the corresponding user's public key from the block chain and encrypts it to obtain c_m. Then, the system encrypts the user attributes using the system private key d_s and obtains the encrypted file c_{att}, which is sent to the user along with the c_m. The user can obtain the system public key e_s from the block chain, and then decrypt the c_{att} information sent by the system. If the decryption results correspond to their own attributes, then the information is sent by the system. The user then decrypts the c_m using his private key, thereby preventing attackers from recommending junk files to the user and preventing attackers from stealing information.

$$Score(M, u) = \frac{\alpha P(M|u) + \beta K(M|u)}{2} \tag{10}$$

V. EXPERIMENT AND ANALYSIS

A. SOCIAL CIRCLE EXTENSION

While extending the community, there are two parameters need to be estimated, which are parameter k in KSM algorithm and threshold δ in CSE algorithm. First, k is assigned as a certain value and δ varies. Then, δ is set constant, and k changes. The experimental process is shown in Figure 4 and 5.

From what can be seen in the result, when $\delta = 0.6$ and $k = 4$, the best performance of the algorithm can be achieved. Thus, δ and k are assigned as above.

In MCRA, EM algorithm is adopted to train the parameters in LDA. After that, the topic number in Target User Topic Model should be estimated based on the recommendation performance. The MAP would be chosen as evaluation criteria. The topic numbers are changed from 10 to 200 by virtue of experience, and the interval is 10. Based on the parameters determined above, CMCR could be implemented.

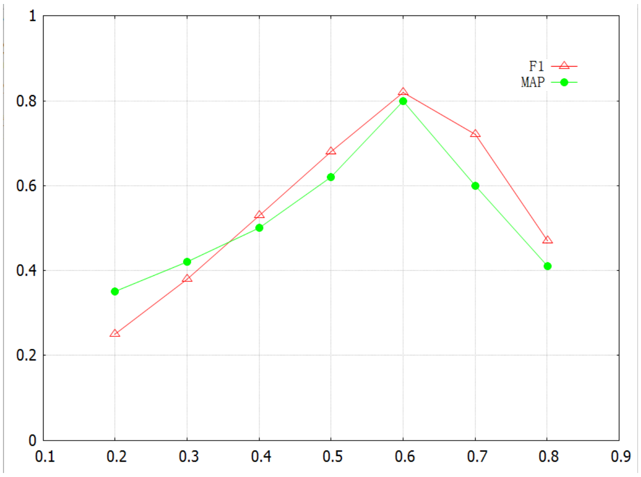


FIGURE 4. Result of changing δ when $k = 4$.

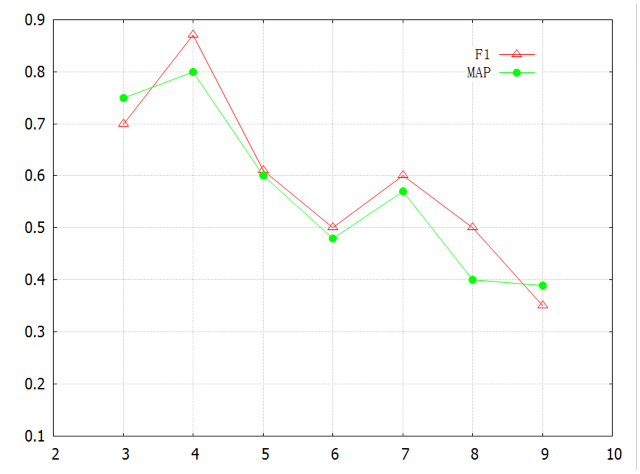


FIGURE 5. Result of changing k when $\delta = 0.6$.

B. USER AUTHENTICATION AND RELATIONSHIP ENCRYPTION MODEL

1) We introduce the authentication with block-chain to verify the validity of the user identity. At the same time, we make use of the acknowledged and non-modifiable of the block chain to store the generated public key on the block chain to bind the block chain address to the public key in order to avoid other people's public key forgery and tampering. Identity can be confirmed only by authenticated users in order to avoid malicious users attacking legitimate users. Also, a user has an address corresponding to the public key, which avoids Sybil attacks to a certain extent.
2) Meanwhile, to prevent honest but curious users take advantage of their interests to get interest from other users, we do not directly transfer relation information in plain text when matching. In this paper, we construct a vector with relations, and make twice hash encryption for relation attributes. Because of the irreversibility of hash encryption, the user only gets the MD5 value of the property so that mismatched users cannot obtain the user relationship attributes. Only same attributes can be obtained for same MD5 to match.
3) Using MD5, the dimension of the string after the first hash encryption can be reduced, then not only the twice hash make encryption more secure to guarantee the uniqueness of attribute encryption, but the number of bits is small and the matching process can increase matching efficiency.
4) The technology of digital signature: in order to guarantee the identity of the user who sends interest, the user should encrypt the data packets with his private key.

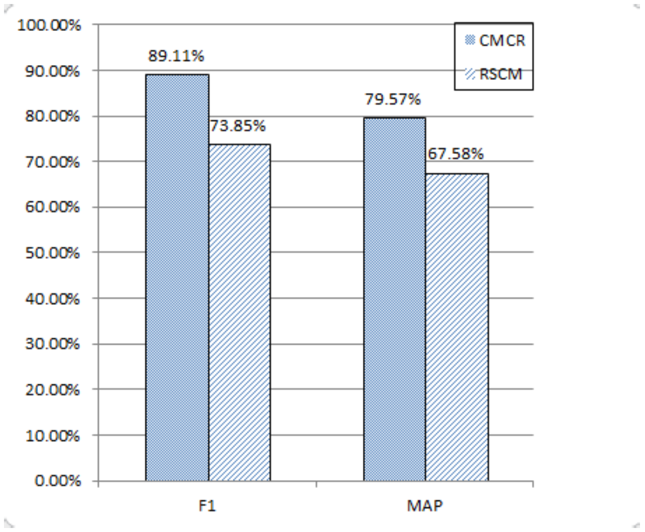


FIGURE 6. Comparison of CMCR and RSCM.

C. THE COMPARISON OF SOCIAL CIRCLE

To verify the efficiency of the expansion, we compare CMCR with RSCM algorithm [18] and K-means algorithm [19]. When the social circle is expanded, K-means algorithm directly uses the number of previously detected communities as the initial number of centers. Moreover, K-means algorithm puts each node into a community, so it is reasonable to use the results of the previous experiment as constraint, ignoring the users who are not included. The result is shown in the Figure 6 and 7.

The results of above experiment show that our algorithm can produce satisfactory results in detecting communities, compared with the RSCM algorithm and K-means algorithm.

D. SECURITY TESTING

Hash encryption algorithm is very effective. There are mainly two common method attacks on the hash encryption: find collision method and exhaustive method.

The first method is finding collision. Different strings would lead to same hash values when a collision occurs. Therefore, the attacker could crack the MD5 by obtaining the same hash value with the one using in the encryption process.

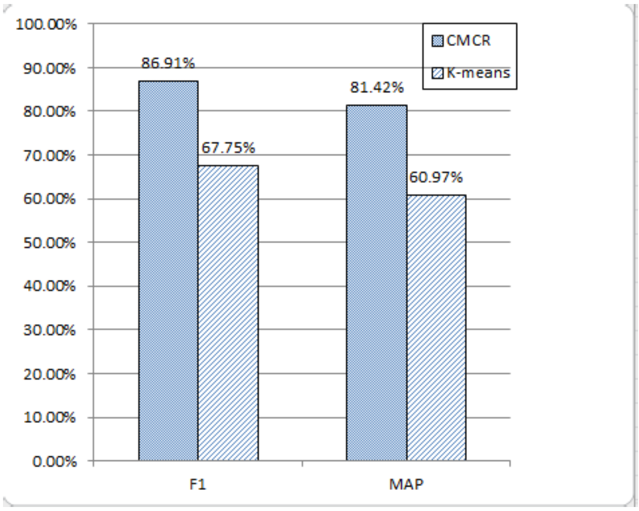


FIGURE 7. Comparison of CMCR and K-means.

TABLE 2. The time and storage of exhaustion.

N_{key}	4	5	6	7	8
T_{key}	0.015s	0.916s	56.8s	58.7min	60.7h
S_{key}	14.1Mb	873.7Mb	52.9Gb	3.2Tb	198.6Tb

But there is no need to worry about this situation, since there is no effective way to find collision method for MD5 and SHA1. So far, the order of magnitude of the most effective crack algorithm to solve MD5 is 2^{69}. But this situation is still limited to theoretical analysis. In fact, 2^{69} is still an impossible number for practical application.

Another method attacking the hash encryption method is exhaustive. For some simple password, this method is very efficient and easily implemented, for example, "123456" and "000000". Because the scanning scope of the exhaustive method is often a single character set, interval with the law, or the combination of the words in the dictionary, so exhaustive method is difficult to work in the cases of complex passwords. The number of characters that may appear at each bit in the hash encryption method is Num_{possi}., the number of bits N_{key} and the time required to crack a password is T_{bit}. The cracked time T_{key} is defined as Formula (11).

$$T_{key} = T_{bit} \times Num_{possi}^{N_{key}} \tag{11}$$

The required storage space S_{key} can be calculated in a similar way. Assuming that Num_{possi} is 62 and T_{bit} is 1ns. As the number of bits N_{key} increases, the cracked time T_{key} and the required storage space S_{key} are shown in Table 2. It is almost impossible to break hash encryption with the ordinary exhaustive method.

The violent crack method can be simplified using a method called a rainbow table. Rainbow table is a precomputed table for breaking the hash value of the password, used in the inverse operation of the encryption hashing function. The rainbow tables are often used to recover the fixed length plain text passwords consisting of the characters in finite set. In brief, this is an effective method cracking of some particular algorithms, especially the asymmetric algorithm, such as MD5 algorithm. Ignoring the time required for the query, the larger the table is, the lower the cost of cracking is. However, for other crack methods, such as collisions, the effect of cracking would be poor. Especially for variable-length keys and other modern advanced algorithms, the effect will be greatly reduced. The use of secondary encryption method, guarantees that the possibility of being cracked is very small using the rainbow table.

Based on the analysis, the brute force is still the main crack hash encryption method. So in order to reinforce the security of encrypted information, we use multiple hash methods. The method is called multi-hash that uses multiple encryptions on the information with the hash method through user-defined Key. If the Key is complicated enough, it's very difficult to crack in exhaustion method. We use Formula (12) as the hash method:

$$R = MD5(SHA1(S)) \tag{12}$$

Even if S is simple, it is still difficult to computing all possibilities in a reasonable time cracking the hash encryption and MD5. This method further ensures the security of the data. The hash value generated by our method is tested by existing online resources.

VI. CONCLUSION

Based on CMCR, we propose a protocol and Authentication with Block-chain algorithm to protect the user privacy information in the social community detection. Considering user with higher closeness, we use authentication mechanism based on the block-chain, and encrypt the relationship with Hash function for better security. Then, we use the text encryption protocol in the text recommendation process to ensure the security of information. Experiments have proven that our improvements have obtained better results.

In the future, we will further improve the security of protocol framework on need basis. We plan to adjust the structure of the block-chain and improve the encryption algorithm to enhance the security of social networks.

REFERENCES

[1] J. Akshay, S. Xiaodan, F. Tim, and T. Belle, "Why we twitter: Understanding microblogging usage and communities," in *Proc. ACM 9th WebKDD 1st SNA-KDD Workshop Web Mining Soc. Netw. Anal.*, San Jose, CA, USA, 2007, pp. 56–65.

[2] Z. Zhao *et al.*, "Topic oriented community detection through social objects and link analysis in social networks," *Knowl.-Based Syst.*, vol. 26, pp. 164–173, Feb. 2012.

[3] R. Yu, "Communities mining and recommendation for large-scale mobile social networks," in *Proc. Int. Conf. Wireless Algorithms, Syst., Appl.*, Cham, Switzerland, 2017, pp. 266–277.

[4] T. Sakuma and R. Sasaki, "Proposal and evaluation of the digital certificate system with sumi-coating module for privacy protection," in *Proc. IEEE Comput. Softw. Appl. Conf. Workshops*, Jul. 2012, pp. 200–205.

[5] J. Duan, Y. Ai, and X. Li, "LDA topic model for microblog recommendation," in *Proc. IEEE 8th Int. Conf. Asian Language Process.*, Suzhou, China, Oct. 2015, pp. 185–188.

[6] X. Deng, G. Li, and M. Dong, "Finding overlapping communities with random walks on line graph and attraction intensity," in *Proc. 10th Int. Conf. Wireless Algorithms, Syst., Appl.*, Qufu, China, 2015, pp. 94–103.

[7] J. Yang and J. Leskovec, "Overlapping community detection at scale: A nonnegative matrix factorization approach," in *Proc. ACM 6th Int. Conf. Web Search Data Mining*, Rome, Italy, 2013, pp. 587–596.

[8] S. Fortunato and C. Castellano, "Community structure in graphs," in *Computational Complexity*. New York, NY, USA: Springer-Verlag, 2012, pp. 490–512.

[9] H. William, C. S. Matthew, B. Paul, and F. S. Nagiza, "On perturbation theory and an algorithm for maximal clique enumeration in uncertain and noisy graphs," in *Proc. ACM 1st ACM SIGKDD Workshop Knowl. Discovery Uncertain Data*, Paris, France, 2009, pp. 48–56.

[10] M. J. Freedman, K. Nissim, and B. Pinkas, "Efficient private matching and set intersection," in *Advances in Cryptology—EUROCRYPT*. Berlin, Germany: Springer-Verlag, 2004, pp. 1–19.

[11] C. Hazay and Y. Lindell, "Efficient protocols for set intersection and pattern matching with security against malicious and covert adversaries," in *Proc. 5th Conf. Theory Cryptogr. (TCC)*, 2008, pp. 155–175.

[12] W. Dong, V. Dave, L. Qiu, and Y. Zhang, "Secure friend discovery in mobile social networks," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 1647–1655.

[13] J. Lei, L. En-tao, and W. Guo-jun, "Privacy-preserving friend matching mechanism in mobile social networks," *J. Chin. Comput. Syst.*, vol. 37, no. 9, pp. 1980–1985, 2016.

[14] M. Nofer *et al.*, "Blockchain," *Bus. Inf. Syst. Eng.*, vol. 59, no. 3, pp. 1–5, 2017.

[15] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," [Online]. Available: https://bitcoin.org/bitcoin.pdf

[16] D. Fu and L. Fang, "Blockchain-based trusted computing in social network," in *Proc. IEEE Int. Conf. Comput. Commun.*, Oct. 2017, pp. 19–22.

[17] R. Ali and A. K. Pal, "A secure and robust three-factor based authentication scheme using RSA cryptosystem," *Int. J. Bus. Data Commun. Netw.*, vol. 13, no. 1, pp. 74–84, 2017.

[18] H. Qin, T. Liu, and Y. Ma, "Mining user's real social circle in microblog," in *Proc. IEEE 4th Int. Conf. Adv. Soc. Netw. Anal. Mining*, Istanbul, Turkey, Aug. 2012, pp. 348–352.

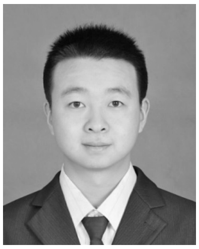
[19] Q. Ba, X. Li, and Z. Bai, "A similarity calculating approach simulated from TFIDF in collaborative filtering recommendation," in *Proc. 5th Int. Conf. Multimedia Inf. Netw. Secur.*, Beijing, China, Nov. 2013, pp. 738–741.

RUIGUO YU received the B.S. degree in computer software and the M.S. and Ph.D. degrees in computer application technology from Tianjin University, China. He is currently an Associate Professor at the School of Computer Science and Technology, Tianjin University. His current research interests include recommended algorithm research and application, text feature extraction and clustering, network information retrieval, and public opinion monitoring. He has participated in a number of projects including Natural Fund projects.

JIANRONG WANG received the B.S., M.S., and Ph.D. degrees in computer application technology from Tianjin University, China. He is currently an Associate Professor at the School of Computer Science and Technology and servers as the Deputy Director of the IT discipline innovation and entrepreneurship training base, Tianjin University. His current research interests include speech recognition, machine learning, human–computer interaction, and other aspects of research work. His main concerns include feature extraction of color image used in speech recognition research, feature extraction of depth information, multichannel audio and video feature fusion and decision fusion, research on machine learning algorithms, including music genre classification and robot self-localization in humming recognition environment.

TIANYI XU received the bachelor's degree from the School of Electronic Engineering and Automation in Tianjin University, China, in 2012, and the master's degree from the School of Computer Science and Technology, Tianjin University, China, in 2015. From 2015, he was an Assistant Engineer with the School of Computer Science and Technology, Tianjin University. His research interests are network and data mining.

JIE GAO received the B.S. degree from the School of RenáI, Tianjin University, China, in 2013 and the M.S. degree from the School of Computer Science and Technology, Tianjin University in 2016. In 2016, she was an Assistant Engineer at the School of Computer Science and Technology, Tianjin University. Her research interests are network and data mining.



YONGLI AN received the B.S. degree from the School of Communication Engineering, Xi'an University of Posts and Telecommunications, China, in 2016. He is currently working toward the M.S. degree at the School of Computer Science and Technology, Tianjin University, China. His research interests are data mining, big data, and social networks.

GONG ZHANG received the B.S. degree from the North China University of Technology of Computer Science and Technology, China, in 2016. She is currently working toward the M.S. degree at the School of Computer Science and Technology, Tianjin University, China. Her research interests are social networks and data mining.

MEI YU received the Ph.D. degree in computer application technology from Tianjin University. She is currently a Professor in Computer Networks, Data Mining, Database at Tianjin University. As the coach of the Tianjin University ACMICPC Team, she led the team to receiving a number of awards at the Asian Regional Contest of the ACM International Collegiate Programming Contest, and reaching the world finals twice. She serves as the instructor of the Tianjin University IT discipline innovation and entrepreneurship training base, responsible for the base construction.