Hindawi
Journal of Electrical and Computer Engineering
Volume 2017, Article ID 3053202, 7 pages
https://doi.org/10.1155/2017/3053202

Research Article

A Variable Weight Privacy-Preserving Algorithm for the Mobile Crowd Sensing Network

Jiezhuo Zhong,[1] Wei Wu,[1,2] Chunjie Cao,[1,3] and Wenlong Feng[1]

[1]*College of Information Science and Technology, Hainan University, Haikou, Hainan, China*
[2]*Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya, Hainan, China*
[3]*State Key Laboratory of Marine Resource Utilization in the South China Sea, Hainan University, Haikou, Hainan, China*

Correspondence should be addressed to Chunjie Cao; chunjie_cao@126.com

Received 6 February 2017; Accepted 3 May 2017; Published 5 June 2017

Academic Editor: Zhuo Lu

Mobile crowd sensing (MCS) network collects scenario, environmental, and individual data within a specific range via the intelligent sensing equipment carried by the mobile users, thus providing social decision-making services. MCS is emerging as a most important sensing paradigm. However, the person-centered sensing itself carries the risk of divulging users' privacy. To address this problem, we proposed a variable weight privacy-preserving algorithm of secure multiparty computation. This algorithm is based on privacy-preserving utility and its effectiveness and feasibility are demonstrated through experiment.

1. Basic Theories

1.1. Architecture of Mobile Crowd Sensing Network. Mobile crowd sensing (MCS) network [1] takes the ordinary mobile terminals as the basis sensing units. The sensing task distribution and sensing data collection are achieved through collaboration via the mobile Internet. This represents a large-scale complex social sensing task. "Crowd" refers to the aspect of mobilizing the power and intelligence of the general public, and "sensing" is the process of acquiring the users' behavioral data under different scenarios using the sensors.

Figure 1 shows a typical MCS framework, which consists of the mobile users and the sensing platform. Mobile users are millions of mobile intelligent terminals, into which sensors are embedded (e.g., GPS, gravity sensor, temperature sensor, camera, microphone, and acceleration sensor). These sensors collect various sensing data, which are updated to the sensing platform via the mobile network or short-range wireless communication network. Upon receiving the data, the sensing platform will commence data analysis and processing. The processed data will be directly applied to a diversity of universal social sensing services. After the data analysis and processing are finished, each parcel of data will be evaluated. The mobile users participating in the sensing tasks will be awarded based on the specific incentive mechanism, so as to attract more users into the large-scale sensing task. Liu proposed schemes based on both the Monopoly and Oligopoly models enhancing the location privacy of MCS applications by reducing the bidding and assignment steps in the MCS cycle [2]. Jin proposed a differentially private incentive mechanism that preserves the privacy of each worker's bid against the other honest-but-curious workers [3]. Furthermore, many researchers focused on the detailed information extraction processing in MCS including Hybrid Deep Learning Architecture [4] and Fog Computing and Data Aggregation Scheme [5, 6].

1.2. Application of the MCS Network. The MCS network comprising the mobile intelligent terminals and the mobile sensors is capable of large-scale, complex, fine-grained, and thorough data sensing and collection. For example, the use of MCS network for the collection, analysis, and fusion of the urban traffic flow information can provide highly efficient and convenient path planning and driver assistance system for the mobile users. The MCS network can also provide the decision-making support for urban transport planning and for the formulation of a safe and highly efficient urban transportation network. The MCS network-based sensing

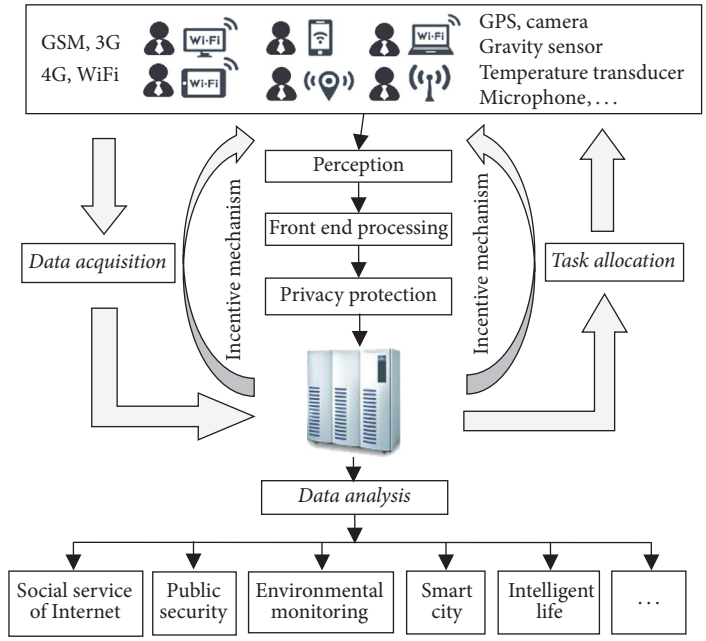

Figure 1: System structure diagram of mobile crowd sensing network.

and monitoring of urban domestic infrastructures offer convenient life services for the local residents. The wide prevalence of the mobile intelligent terminals is a solid guarantee for the high-efficiency and low-cost and large-scale monitoring of natural environment in the cities.

2. Privacy Protection Mechanism for MCS Network Users

2.1. Privacy Preserving in MCS Network. The sensing data collected by the MCS network are largely user privacy. Location data usually contain the sensitive information such as users' address, scope of activity, and transportation route. The mining of users' state of motion can obtain the sensitive information of users' living habits and health conditions. The biological data collected contains the information of users' voice, fingerprints, and basic physiological characteristics. The routine usage data of the mobile intelligent terminals are associated with the user privacy of a deeper level, including the users' hobbies and behavioral traits. Once the user privacy is divulged, there may be violation of privacy, harassment, fraud, or even direct economic loss. Therefore, designing the data security architecture for dynamic privacy protection under the MCS network is an urgent issue.

2.2. Related Technology of Privacy Preserving. The major privacy-preserving techniques used for MCS network are divided into the following types.

(1) Generalized Privacy-Preserving Algorithm. Anonymization is performed while sharing the sensing data, so that the sensitive information about the user's identity is removed without harming the meaningful deduction based on the anonymized sensing data. However, the currently used anonymization methods are usually greedy algorithms which have low execution efficiency.

(2) Perturbation-Based Privacy-Preserving Algorithm. The raw sensing data are perturbed by adding a random number, noises, and exchanges, so that the other party cannot mine the raw sensing data and privacy policies. The main difficulty with data perturbation is how to strike a balance between data correctness, privacy, and security.

(3) Secure Multiparty Computation (SMPC). This technique integrates data encryption and multiple parties are involved in the computation and mining. Because none of the parties have access to complete data, the users' privacy can be ensured. SMPC is now used for collaborative computing among a group of untrusted parties. Many researches have been carried out over the SMPC problem. In 2000, Lindell proposed the method of secure multiparty decision tree (ID3) to protect the data privacy of users [7]. Asharov proposed the threshold homomorphic encryption scheme to improve efficiency of the privacy protection algorithm [8]. In 2014, the threshold-based encryption of K-means outsourcing computing proposed by Liu is a more efficient privacy protection algorithm [9].

Proper application of information technology and algorithm design are the two major concerns in privacy protection. However, the users' attitudes towards privacy are generally neglected. A survey [10] indicates that 17% of the Internet users are still unwilling to provide their authentic information even under privacy protection; 56% of the Internet users are more willing to provide their authentic information in the presence of proper privacy protection; the remaining 27% of the Internet users do not particularly care about their privacy

and will provide the authentic information with or without privacy protection. It is obvious that the users' attitudes towards privacy affect their willingness to share the personal information. Users may react differently to the prospect of disclosing different personal information. But under some incentive mechanisms, the psychological response of the users to the disclosure of different sensitive information may vary.

This study constructed an MCS network-based privacy-preserving algorithm by reference to SMPC. The weight function of privacy preference was built by combining the analysis of the users' sensitivity to the disclosure of privacy and classification of the privacy level of the sensing data. This proposed algorithm can effectively prevent the divulging of privacy information while achieving a maximal acquisition and analysis of the sensing data.

3. Variable Weight Privacy-Preserving Algorithm

3.1. Measure of User Privacy Sensing

3.1.1. User Multiattribute Assumption. Suppose there exists Euclidean space, in which n dimensions represent n solutions to one problem; f_j denotes the attribute j, and G is a set of attributes, $G = \{f_1, f_2, \ldots, f_n\}$. x_i denotes one solution, and X is a set of one solution, $X = \{x_1, x_2, \ldots, x_m\}$. $x_{ij} = f_j(x_i)$ denotes the attribute value of the solution under attribute f_j. $D = (x_{ij})_{m\times n}$ denotes a decision-making matrix of solution X under the attribute G:

$$D = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix}. \tag{1}$$

Considering the varying sensitivity to privacy, the users show different willingness to share their privacy in the MCS network. The influence factors of this willingness are divided into profit factors and risk factors, each of which is measured differently. Let $M = \{1, 2, \ldots, m\}$ be the set of the profit attributes, and $N = \{1, 2, \ldots, n\}$ be the set of risk attributes. The two sets are normalized by multiple attribute decision-making using the following formula:

$$y_{ij} = \frac{\left(x_{ij} - \min_i x_{ij}\right)}{\left(\max_i x_{ij} - \min_i x_{ij}\right)} \quad i \in M,\ j \in N. \tag{2}$$

After the transform, the synthetic matrix is $Y = (y_{ij})_{m\times n}$.

$$Y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{pmatrix}. \tag{3}$$

3.1.2. Weight Determination of Privacy Perception Attributes. As the users differ in privacy perception, each attribute will carry the information of different user preferences. Therefore, the given user preference can be expressed as the weight of the individual, and the weight of each attribute is expressed as

$$W = (w_1, w_2, \ldots, w_n)^T, \quad \sum_{j=1}^{n} w_j = 1. \tag{4}$$

The utility of each user is expressed as the sum of the weighted attributes. Hence, the user utility U_i is

$$U_i = \sum_{j=1}^{n} y_{ij} \cdot w_j, \quad (j = 1, 2, \ldots, n). \tag{5}$$

The utility analysis of users' privacy perception will provide not only the weight parameters for the SMPC, but also some suggestions for the collection modes of the sensing data under the MCS network. For example, the privacy information sensitive to most users will be prevented and a reasonable incentive mechanism can be designed on this basis. This is very important for increasing the confidence and participation level of users with a lower utility of privacy perception.

3.2. Variable Weight SMPC-Based Privacy-Preserving Algorithm

3.2.1. SMPC-Based Algorithm. SMPC can be conceptualized by the following mathematical model: n participants $P_1, P_2, \ldots, P_n$ of the protocol jointly implement the function $f(x_1, x_2, \ldots, x_m)$. $S_{\text{input}} = \{x_1, x_2, \ldots, x_m\}$ is the set of input variables. The set of input variables S_{P_i} provided by the participant P_i $(i \in \{1, 2, \ldots, n\})$ is a subset of S_{input}, which satisfies $\bigcup_{P_i} S_{P_i} = S_{\text{input}}$, $S_{P_i} \cap S_{P_j} = \phi$ $(i \neq j)$. It is required in the computing of the function that the input S_{P_i} from any participant P_i $(i \in \{1, 2, \ldots, n\})$ is not known to other participants P_j $(j \neq i)$.

The essence of SMPC is a data encryption algorithm using the encryption scheme so as to ensure data privacy. Rivest et al. [11] proposed the concept of fully homomorphic encryption in 1978, aiming to construct an encryption mechanism that supports ciphertext retrieval. Goldwasser [12] studied the strategies used by mobile attackers in the secure channel model. They generalized the threshold mechanism to the ordinary SMPC. The plaintext will be revealed only when at least t participants are involved in the collaborative decryption. This effectively restricts the access to the final SMPC output and the participants will not disclose the data.

3.2.2. Weighted Threshold Secret Sharing Scheme Based on Mignotte Sequence. The weighted threshold secret sharing scheme refers to that each participant assumes a different role, based on which different weights are assigned. The conventional weighted threshold secret sharing schemes achieve only works on the premise of assigning more secret shares

to those who are given special permission. However, this will increase the insecurity of key management and transmission. In this study, we adopted the weighted threshold secret sharing scheme based on *Mignotte* sequence. Regardless of the weight, each participant is only allowed one private key and there is no transmission of secret information between the participants and the dealer. Therefore, the cost of key transmission and storage is spared.

Mignotte sequence is defined as follows [13]:

Let $k, n \in Z,\ n \geqq 2,\ 2 \leqq t \leqq n$. If the integer sequence $m_1, m_2, \ldots, m_n$ satisfies

(1) $m_1 < m_2 < \cdots < m_n$;

(2) $(m_i, m_j) = 1$, where $1 \leqq i < j \leqq n$;

(3) $\prod_{i=0}^{t-2} m_{n-i} < \prod_{i=1}^{t} m_i$,

then sequence $m_1, m_2, \ldots, m_n$ is called a (t, n)-*Mignotte* sequence.

The weighted threshold secret sharing scheme based on the *Mignotte* sequence is designed.

(1) Parameter Configuration. In this scheme, the dealer assigns the weights to each participant using a digit with a length of large prime. The secret to be shared is determined and the relevant system parameters are configured. There are n participants and they constitute the set $U = \{u_1, u_2, \ldots, u_n\}$. The weight vectors of the participants are correspondingly $W = (w_1, w_2, \ldots, w_n)$. The threshold is t, and the secret to be shared is s.

(2) Construction and Expansion of Mignotte Sequence. The dealer needs the system parameters to construct an expanded Mignotte sequence fit for the weighted threshold secret sharing scheme. Meanwhile, the converted scheme should be equivalent to the original scheme. A (t, n)-*Mignotte* sequence is constructed as $m_1', m_2', \ldots, m_n'$, which is expanded into

$$\overbrace{m_1', \ldots, m_1'}^{w_1}, \overbrace{m_2', \ldots, m_2'}^{w_2}, \ldots, \overbrace{m_n', \ldots, m_n'}^{w_n}. \tag{6}$$

The above sequence is a sequence of n' primers, where $n' = \sum_{i=1}^{n} w_i$ which makes the sequence satisfy the following conditions:

(a) The product β of the last $t-1$ numbers is smaller than the product α of the first t numbers.

(b) $\beta < s < \alpha$.

Let $m_1 = (m_1')^{w_1}, m_2 = (m_2')^{w_2}, \ldots, m_n = (m_n')^{w_n}$.

From above, it can be known that sequence $m_1, m_2, \ldots, m_n$ has the following property: $(m_i, m_j) = 1$ $(1 \leqq i \leqq j \leqq n)$.

When $w_n + w_{n-1} + \cdots + w_j < t < w_1 + w_2 + \cdots + w_i$, $m_n m_{n-1} \cdots m_j < s < m_1 m_2 m_i$, $(1 \leqq i \leqq j \leqq n)$.

Thus sequence $m_1, m_2, \ldots, m_n$ is the expanded Mignotte sequence, denoted as (W, t, n)-*Mignotte* sequence. This sequence is revealed.

(3) Generation of Secret Shares. The dealer computes the secret shares of each participant according to the *Mignotte* sequence $m_1, m_2, \ldots, m_n$ and the shared secret:

$$\begin{aligned} s_1 &= s \bmod m_1, \\ s_2 &= s \bmod m_2, \\ &\vdots \\ s_n &= s \bmod m_n. \end{aligned} \tag{7}$$

This S_i is sent to the participant u_i via the secret channel.

(4) Secret Restoration. Suppose there are k participants who constitute the set A, $A = (u_1, u_2, \ldots, u_k)$, and restore the secret. The vector weights for each participant in A constitute the set $W = (w_1, w_2, \ldots, w_n)$.

When the sum of the weight vectors of each participant in A is above or equal to the threshold, that is, $\sum_{i=1}^{t} w_i \geq t$, the following congruence equations are constructed:

$$\begin{aligned} x &= s_1 \bmod m_1, \\ x &= s_2 \bmod m_2, \\ &\vdots \\ x &= s_k \bmod m_k. \end{aligned} \tag{8}$$

$x = \sum_{i=1}^{k} s_i M_i^{-1} M_i (\bmod m)$, where $M_i = m \mid m_i$, $M_i^{-1} M_i \equiv 1 (\bmod m_i)$, $(1 \leq i \leq k)$ and the solution x is the shared secret s.

4. Implementation and Deployment

MapReduce System was used for the high-efficiency parallel processing in the large-scale matrix multiplication in the weighted threshold secret sharing scheme. On the simple data center comprising 5 host machines, the Hadoop distributed storage and computing environment was deployed as a mimic of the sensing platform in the MCS network. One host machine was the Master node, which was deployed with the roles of NameNode and JobTracker for the management of distributed data and task decomposition; 4 host machines were the Slaver and were deployed with the roles of DataNode and TaskTracker for the distributed data storage and task execution. The implementation and deployment (Figure 2) are illustrated below.

(1) The initialization program at the data center would preset the system parameters. The threshold t was determined. The weight vectors of each participant were initialized. The key management system as the trusted third party generated n pairs of homomorphic public and private keys. The public keys hom_PK were the same, and the n private keys were distributed to different participant nodes.

(2) The block function MR_Splitter( ) in the MapReduce System was responsible for dividing the sensing data files submitted by the clients in the MCS network into blocks. Each block was 64 M. The data blocks were encrypted using the

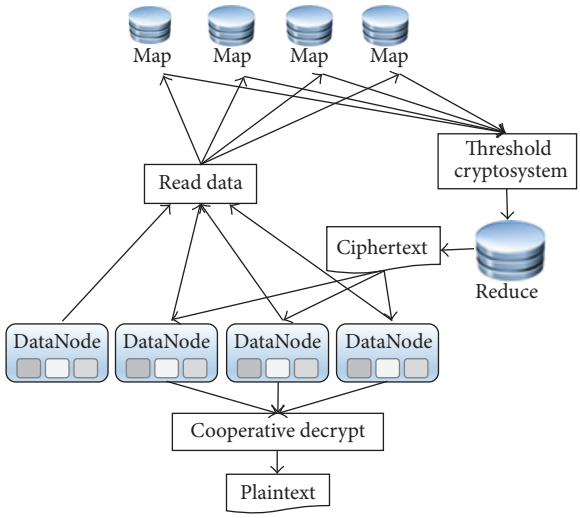

Figure 2: Flowchart of deployment.

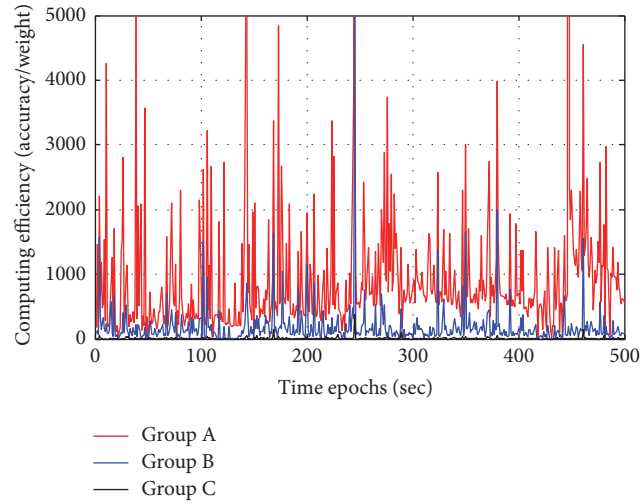

Figure 3: Privacy computing efficiency iteration results of all groups.

public key hom_PK. The encrypted data block file is stored in the distributed file system of the DataNode.

(3) An intermediate ⟨key, value⟩ pair was computed during the matrix multiplication in the privacy-preserving algorithm. The map nodes were allocated to each operation. Before the mapper output the ⟨key, value⟩ pair, the ciphertext for each participant was generated using formulae (7).

(4) Reducer replicated the intermediate output of the corresponding division from the mapper output terminal to the local file system.

(5) At least t participants were involved in the decryption of the ciphertext using the decryption algorithm in formulae (8). These participants would share the decrypted information with other participants. The information decrypted by the t participants was then combined with the information decrypted by the remaining participants to obtain the final result.

For users in network society, we divide them into three groups according to the weights aforementioned in Sections 2 and 3, that is, privacy careless person (group A), practical privacy person (group B), and the group who protect their privacy strictly (group C). In group A, they are not so sensitive with privacy and willing to share their true information. In group B, they may share personal files while policies and regulations are carefully learned. In group C, they are not interested in any sharing information activities at any circumstances.

In a certain survey, the percentage results of groups A, B, and C are obtained as 33.1%, 57.4%, and 9.5% from 352 users on the Internet, and we can initiate the weights of sharing by 0.9, 0.5, and 0.1. These parameters are easily adjusted during privacy protection mechanism proposed here.

As is shown in Figure 3, the three groups in privacy iteration results are given. Group A indicates that since they are not concerned about their information, those provided data are true and the efficiency is acceptable. Group B is matching data on the condition that they believe the privacy is protected, so that their efficiency is not stable and high. Group C are not willing to share their information, and their provided information is not all correct, which influences the computing fundamentally.

5. Algorithm Performance Analysis

5.1. Security Analysis. The private information of each participant is randomly divided into m fragments in a certain way. Each participant selects one fragment randomly and preserves it. The remaining fragments are randomly allocated to other participants. After the fragments are reallocated according to the protocol, each participant will own an equal amount of fragments. Each participant owns one fragment of his or her information plus one fragment transmitted from another participant. Therefore, even if participants P_{i-1} and P_{i+1} conspire, they can only infer the reallocated information of participant P_i and do not know other private information N_i. Any two conspiring participants can only infer the reallocated information of the third party. Then, combining with the information fragments owned by themselves, they can infer the private information of the third party. But when there are more than 3 participants, it will be very difficult to infer all information of the other participants by conspiracy. When there are more than 4 participants and when most participants are honest, the possibility of information leak will approach 0.

5.2. Complexity Analysis. Computational complexity: each round of computation consists of m operations (different from the m aforementioned), and m rounds involve m^2 operations. Thus the computational complexity $S(m)$ is expressed as $S(m) = m^2$, as shown in Figure 4.

Communication Complexity. Each participant needs to transmit $m - 1$ fragments to other participants. Therefore, in the fragment transmission stage, $m(m - 1)$ communications will occur. In the computing stage, each participant needs

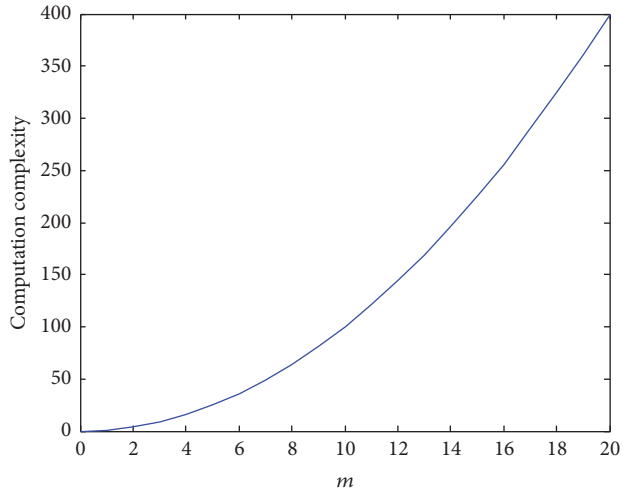

Figure 4: Computation complexity.

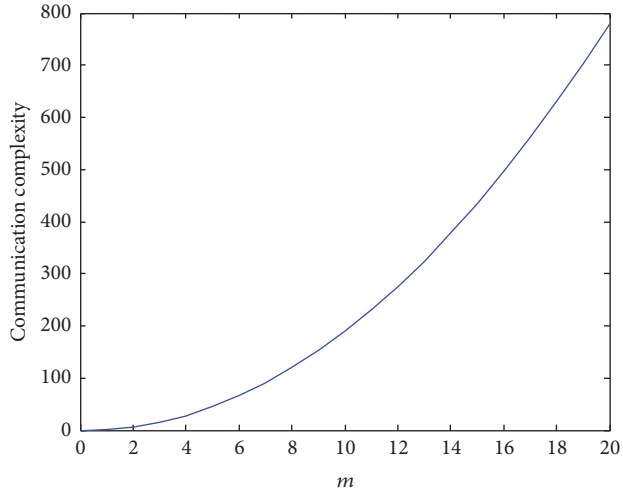

Figure 5: Communication complexity.

to transmit the summation of some fragments to other participants over the ring structure. Therefore, each round of computation consists of m communications, and m rounds involve m^2 communications. The overall communication complexity $C(m)$ of the algorithm is expressed as $C(m) = m(m-1) + m^2 = 2m^2 - m$, as shown in Figure 5.

6. Summary and Forecast

To protect against privacy violation in the MCS network, we proposed a variable weight SMPC-based privacy-preserving algorithm. The weighted threshold secret sharing scheme based on Mignotte sequence was applied for the encryption of the sensing data and private key management. Considering the different attitudes of users towards the disclosure of the private information, the privacy of the information was graded. Thus the weight parameters of the privacy-preserving algorithm were determined based on the utility analysis of the users' privacy perception. The proposed model was deployed in the Hadoop distributed environment to verify its effectiveness and validity. The implementation of the SMCP protocol requires several participants, among which communications are necessary. This will incur significant communication and computational costs. How to enhance the reliability of channel communication and to increase the efficiency of sensing data encryption are issues awaiting resolution.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper and confirm that the mentioned received funding in the Acknowledgments did not lead to any conflicts of interest regarding the publication of this manuscript.

Acknowledgments

This work was supported by the Natural Science Foundation of Hainan Province (no. 20166216 and no. 617033) and Education and Reaching Research Project of Hainan University (no. hdjy1325) investigated by Jiezhuo Zhong; National Natural Science Foundation of China (no. 61661019), the Major Science and Technology Project of Hainan Province (no. ZDKJ2016015), the Natural Science Foundation of Hainan Province (no. 20156217), and the Higher Education Reform Key Project of Hainan Province (no. Hnjg2017ZD-1) by Chunjie Cao; National Science and Technology Support Program (no. 2015 BAH55F01-5) and Natural Science Foundation of Hainan Province (no. 614232) investigated by Wenlong Feng.

References

[1] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: current state and future challenges," *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32–39, 2011.

[2] B. Liu, W. Zhou, T. Zhu et al., "Invisible hand: a privacy preserving mobile crowd sensing framework based on economic models," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 1-1, 2017.

[3] H. Jin, L. Su, B. Ding et al., "Enabling privacy-preserving incentives for mobile crowd sensing systems," in *Proceedings of the IEEE 36th International Conference on Distributed Computing Systems (ICDCS '16)*, pp. 344–353, Nara, Japan, June 2016.

[4] S. A. Ossia, A. S. Shamsabadi, A. Taheri et al., "A Hybrid Deep Learning Architecture for Privacy-Preserving Mobile Analytics," https://arxiv.org/abs/1703.02952.

[5] S. Basudan, X. Lin, and K. Sankaranarayanan, "A privacy-preserving vehicular crowdsensing based road surface condition monitoring system using fog computing," *IEEE Internet of Things Journal*, no. 99, pp. 1-1, 2017.

[6] C. Xu, R. Lu, H. Wang, L. Zhu, and C. Huang, "PAVS: a new privacy-preserving data aggregation scheme for vehicle sensing systems," *Sensors*, vol. 17, no. 3, p. 500, 2017.

[7] Y. Lindell and B. Pinkas, "Privacy preserving data mining," in *Advances in Cryptology—CRYPTO 2000, 20th Annual International Cryptology Conference, Santa Barbara, California, USA,*

August 20–24, 2000, vol. 1880 of *Lecture Notes in Computer Science*, pp. 36–54, 2000.

[8] G. Asharov, A. Jain, A. López-Alt, E. Tromer, V. Vaikuntanathan, and D. Wichs, "Multiparty computation with low communication, computation and interaction via threshold FHE," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7237, pp. 483–501, 2012.

[9] L. Liu and M. Tamer Özsu, *Encyclopedia of Database Systems*, Springer, New York, NY, USA, 2017.

[10] D.-H. Shin, "The effects of trust, security and privacy in social networking: A security-based approach to understand the pattern of adoption," *Interacting with Computers*, vol. 22, no. 5, pp. 428–438, 2010.

[11] R. L. Rivest, L. Adleman, and M. L. Dertouzos, *On Data Banks And Privacy Homomorphism Proc of Foundations of Secure Computation*, Academic Press, New York, NY, USA, 1978.

[12] S. Goldwasser, "Multi party computations: past and present," in *Proceedings of the sixteenth annual symposium on Principles of distributed computing (ACM '97)*, pp. 1–6, August 1997.

[13] M. Mignotte, "How to share a secret," *Lecture Notes in Computer Science*, vol. 149, no. 2, pp. 371–375, 1983.