Computer Communications 119 (2018) 167–178

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Privacy-preserving scheme in social participatory sensing based on Secure Multi-party Cooperation

Ye Tian[a], Xiong Li*[,b], Arun Kumar Sangaiah[c], Edith Ngai[d], Zheng Song[e], Lanshan Zhang[f], Wendong Wang*[,a]

[a] The State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, 100876, China
[b] School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China
[c] School of Computer science and Engineering, VIT University, Vellore-632014, Tamilnadu, India
[d] The Department of Information Technology, Uppsala University, Uppsala, Sweden
[e] Virginia Tech University, Virginia, US
[f] Beijing Key Laboratory of Network System and Network Culture, Beijing University of Posts and Telecommunications, Beijing, 100876, China

ARTICLE INFO

Keywords:
Social participatory sensing
Privacy-preserving
Secure Multi-party Cooperation

2010 MSC:
00–01
99–00

ABSTRACT

Social participant sensing has been widely used to collect location related sensory data for various applications. In order to improve the Quality of Information (QoI) of the collected data with constrained budget, the application server needs to coordinate participants with different data collection capabilities and various incentive requirements. However, existing participant coordination methods either require participants to reveal their trajectories to the server which causes privacy leakage, or tradeoff the location accuracy of participants for privacy, thereby leading to lower QoI. In this paper, we propose a privacy-preserving scheme, which allows application server to provide quasi-optimal QoI for social sensing tasks without knowing participants' trajectories and identity. More specifically, we first suggest a Secure Multi-party Cooperation (SMC) based approach to evaluate participant's contribution in terms of QoI without disclosing each individual's trajectory. Second, a fuzzy decision based approach which aims to finely balance data utility gain, incentive budget and inferable privacy protection ability is adopted to coordinate participant in an incremental way. Third, sensory data and incentive are encrypted and then transferred along with participant-chain in perturbed way to protect user privacy throughout the data uploading and incentive distribution procedure. Simulation results show that our proposed method can efficiently select appropriate participants to achieve better QoI than other methods, and can protect each participant's privacy effectively.

1. Introduction

The ubiquity of various sensors within smart devices has inspired a new wave of research towards social participatory sensing, which was first discussed in [1]. As a people-centric and spatial-based sensing task, social participatory sensing fully utilizes the idea of crowdsourcing [2]. The major difference between social participatory sensing and traditional sensing lies in the fact that, each participant being regarded as a sensor, called social sensor, sensing the surrounding environment to upload data [3]. A group of mobile users subscribe to an application server and a number of task publishers who publish to the application server both task requirements and corresponding incentive budgets.

Quality-of-information (QoI) is a widely used index to describe publisher's requirement, or evaluate the actual performance of social participatory sensing tasks. Broadly speaking, QoI relates to the ability to judge whether information is fit for use for a particular purpose [2,4,5]. Actually, QoI is characterized by a number of attributes, including sensing locations, sensing time period, required amount of data at each location. Above all, coverage rate and redundancy of sensory data in sensing areas are two key attributes in sensing tasks. Low coverage or high redundancy will lead to deficiency of valid data, and finally affects the performance of social participatory sensing task. An evenly-distributed sensory data with good balance in data coverage and redundancy is of great importance to QoI, thus borrowing the concept which was proposed in [2], we adopt *data utility* to evaluate QoI (detail definition please refer to Section 4.2). To achieve better QoI for sensing tasks under budget constraints, application server needs to coordinate appropriate participants for data collection. Existing approaches

* Corresponding author.
E-mail addresses: yetian@bupt.edu.cn (Y. Tian), lixiongzhq@163.com, lixiong@hnust.edu.cn (X. Li), edith.ngai@it.uu.se (E. Ngai), zls326@sina.com (L. Zhang), wdwang@bupt.edu.cn (W. Wang).

http://dx.doi.org/10.1016/j.comcom.2017.10.007
Received 15 January 2017; Received in revised form 12 September 2017; Accepted 7 October 2017
Available online 16 October 2017
0140-3664/

assume that the application server knows the exact locations of all potential participants as a prior condition, it selects a portion of participants to collect more uniformly-distributed sensing data within the incentive budget constraint, avoiding redundant data. Besides, the data tagged with locations are required to be connected with its collector, so that the application server can evaluate the contributions of the selected participants and reward them accordingly.

However, potential privacy disclosure extends far beyond the temptation of incentives, which may prevent part of people from joining social participatory sensing tasks [1]. The key steps in traditional social participatory sensing scenarios [2] can be summarized in five steps: (1) Application server first publishes sensing task with detail requirements, including task area, task time period, required amount of data in each region and incentive budget; (2) Mobile users report their trajectory and requested incentive for each piece of data; (3) Application server selects optimal users as participants according to their data collection capability (how many regions can be covered by his/her trajectory) and requested incentive; (4) Participants report their data with location tag and user ID; (5) Application server evaluates the gathered data and distributes incentives to each participant.

Conspicuously, privacy disclosure and security risks occur in the above steps. First, users' trajectory information may leak in the following aspect: (1) at participant selection stage, application server needs mobile users' trajectory information to compare their coverage on task area, for selecting optimal candidate with best data collecting capability. (2) application server or other third-party server keeps mobile users' ID and IP address for data uploading or incentive distribution. (3) reported sensing data is tagged with location and user ID for incentive distribution. Second, besides the trajectory information, users' sensitive identity information (such as gender, age, income, political tendency and etc.) may leak. An adversary (for example, the application server) can obtain background knowledge from the pieces of reported sensory data (especially for those semantically-rich data, like photo, video and etc), and identify with high confidence the sensitive value of an individual through association rules based background knowledge attack. Third, at incentive distribution stage, it may involve an important security issue, i.e., incentives may be misappropriated by malicious users. Based on the above analysis, the privacy issues are classified into two categories: namely ***visible privacy leakage*** and ***inferable privacy leakage***. More concretely, visible privacy refers to the visible sensitive information which relates to individual's location or trajectory privacy. Correspondingly, inferable privacy refers to those sensitive information about participant's identity which are deduced by adversary through background knowledge attack [6,7].

There has been two kinds of approaches to resolve the conflict between the server's requirement of knowing participants' locations and the participants' requirements of keeping their location private. The first approach assumes that there is a trustful third party (TTP) server, which is responsible for connecting locations and identifications [8,9]. However, this approach relies too much on the TTP as argued by recent approaches [10,11]. Since the TTP knows too much sensitive information of users, it may become the single target of attacks easily. Therefore, most recent solutions are based on the second approach. The main idea is to tradeoff the location accuracy of uploaded data for location privacy. K-anonymity is a representative approach which guarantees that a user is indistinguishable from at least $k-1$ other users, and widely used in privacy of social network [12,13]. To achieve k-anonymity, a location-based query is submitted to server via a centralized location anonymizer, which enlarges the queried location into a bigger region, geographically covering at least $k-1$ other users [14,15]. However, not knowing the accurate location of uploaded data may affect the coordination phase and the incentive distribution phase, and cause redundant data collection or misjudgement of participants' uploaded data.

For the concern of inferable privacy leakage, differential privacy is an emerging technology to provide means for maximizing the accuracy of queries from statistical databases while minimizing the chances of identifying records. Differential privacy is most used in situations when a trusted party holds a dataset of sensitive information (e.g., transaction records, medical records, voter registration information, and etc.) with the goal of providing global, statistical information about the data publicly available, while preserving the privacy of users whose information the data set contains. Unlike the situation mentioned above, the objective of application server in participatory sensing is not for providing public accessible data while preserving participants's sensitive information. On the contrary, the application server itself is not a completely trusted party in participants's eyes. So differential privacy, which is designed for providing secure data release mechanism does not fit well of the scenario that participatory sensing focus on.

Motivated by the application scenario proposed in[2], we first propose a privacy-preserving participant selection approach based on Secure Multi-party Cooperation. The basic concept behind such scheme is to replace centralized computation involving participants' sensitive information with distributed cooperation among participants. On the application server side, it iteratively selects participants according to processed non-sensitive data instead of raw location related data, and finally constructs a participants-chain. On participants side, participants jointly compute their own contribution for a participatory sensing task while keeping each one's location and identity information private. In addition, a distributed mechanism for data aggregation and incentive distribution is also designed based on the constructed participants-chain. The proposed scheme can achieve quasi-optimal QoI for sensing task, guarantee the robustness of data collection, and above all, preserve participants' both location privacy and identity privacy. The major contribution of our work is four-fold:

- As far as we know, the proposed privacy-preserving scheme is the first to address both visible privacy and inferable privacy problems of participatory sensing task.
- We propose a secure multi-party approach to calculate participants' data sensing ability cooperatively among candidates, which provides essential decision-making basis for participants selection while keeping visible privacy private.
- We propose a multi-criteria ranking based participant selection algorithm to achieve quasi-optimal quality of sensing task. Participants are selected iteratively by explicitly considering their data sensing ability, incentive requirements and impact on inferable privacy preservation ability.
- We design a distributed mechanism to support data aggregation and incentive distribution that works with the proposed privacy-preserving participant coordination method.

The rest of this paper is organized as follow: Section 2 reviews the related literatures. Section 3 establishes the architecture. Section 4 elaborates the proposed participant selection mechanism based on Secure Multi-party Cooperation and fuzzy multi-criteria ranking. Section 5 analyzes the supporting mechanism for data aggregation and incentive distribution. Section 6 conducts privacy analysis and evaluates the performance of the proposed scheme by simulations using real mobility traces. Finally, Section 7 concludes the paper.

2. RELATED WORK

Privacy-preserving is an important issue in many systems, such as in cloud computing environment, Fu et al. [16] and Xia et al. [17] proposed two efficient privacy-preserving search schemes over encrypted outsourced data, respectively. Wang et al. [35] presented an agent-based model of manipulating prices in finacial markets through spoofing, which provided way to quantify the effect of spoofing on trading behavior and efficiency. Besides the tradeoff between the quality requirement of sensing task and the budget constraint of incentives in most existing works, as Krumm [18] discussed in their work,

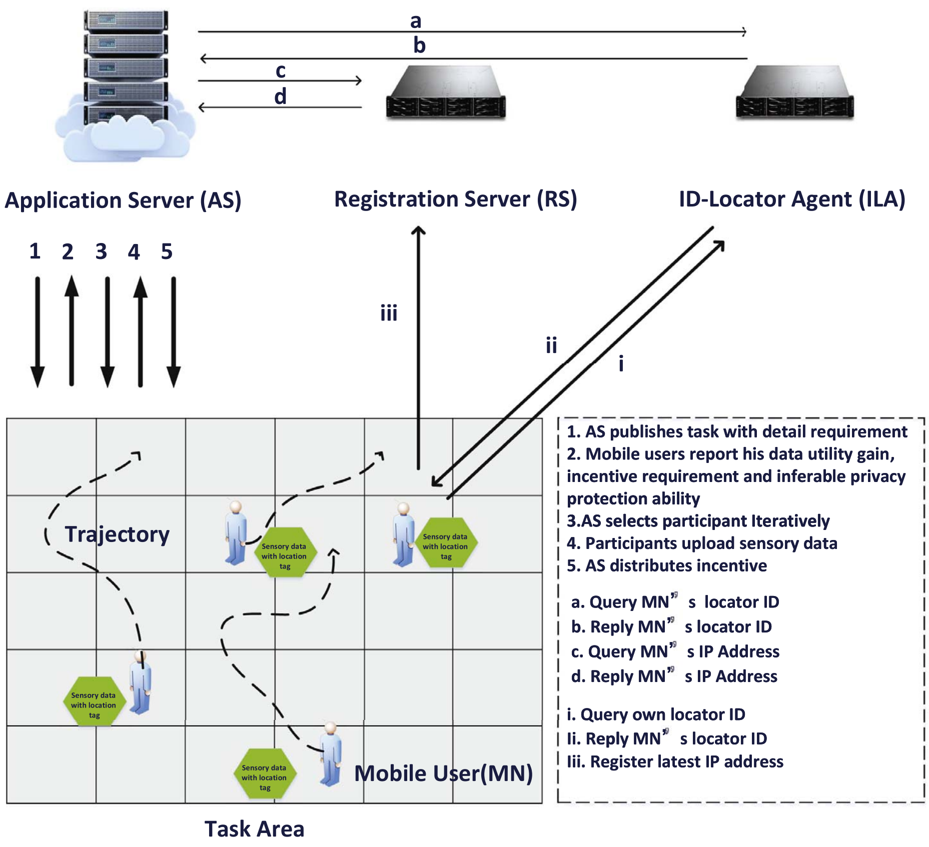

a
b
c
d
Application Server (AS)
Registration Server (RS)
ID-Locator Agent (ILA)
1 2 3 4 5
iii
ii
i
Trajectory
Sensory data with location tag
Mobile User(MN)
Task Area
1. AS publishes task with detail requirement
2. Mobile users report his data utility gain, incentive requirement and inferable privacy protection ability
3.AS selects participant Iteratively
4. Participants upload sensory data
5. AS distributes incentive
a. Query MN’ s locator ID
b. Reply MN’ s locator ID
c. Query MN’ s IP Address
d. Reply MN’ s IP Address
i. Query own locator ID
Ii. Reply MN’ s locator ID
Iii. Register latest IP address

Fig. 1. Application Scenario.

privacy-preserving is also a thorny problem in participatory sensing. Disclosing of participants privacy keeps them away from joining in, especially for those who have strong self-protection awareness, which would cause the enormous power of data contribution being cut down sharply. Existing researches try to propose different approaches to prevent privacy leakage from application server. Rahman et al. [19] reviewed these approaches and categorized them into six categories: pseudonymity, encryption-based, exchange-based, perturbation, sensitive location hiding, and cloaking.

Gao et al. [20] utilized the pseudonymity method in their mix-zones model to hide the association between each participant and his uploaded sensory data. Cristofaro and Soriente [21] proposed an encryption-based method with access control to protect the location privacy of both data producers and consumers. These works rely heavily on a trustful third party, which is easily targeted by attackers [22,23]. Exchange based method, as Boutsis and Kalogeraki [24] discussed in their work, is used to hide the exact trajectory of each participant by exchanging the sensory data of two participants before uploading. This method does not suffer from single point of failure, but application server may fail to evaluate the contribution of each participant. In order to hide the privacy within the sensory data, Ganti et al. [25] relied on the data perturbation method to distort the sensory data with artificial noise. Mun et al. [26] also designed a sensitive location hiding method to generate believable proxy traces instead of the real locations of participants. Cloaking, which has been widely adopted, hides the accurate location of sensory data in a cloaked region as Shin et al. [10] and Wang et al. [11] showed in their works. These works are based on the tradeoff between the accuracy and privacy of uploaded sensory data, and always lead to a loss of quality requirement of sensing task. In contrast to these approaches which protect the privacy of participants during the frequent data communication, we realize that the serious imbalance of data proportion of the whole uploaded sensory data is also likely to expose the most contributor from the masses of participants. In fact, researches that came very close to this concern is privacy preserving mechanisms for background knowledge attack. Social network applications usually contain sensitive and private information about individuals such as relationships, user profile and etc. Some third parties may exploit various kinds of approaches to analyze these data, so the user privacy is likely to be compromised[27–29]. For example, attacker may identify a victim based on the context knowledge such as its degree and neighborhood. To avoid this kind of background knowledge attack, Lin Yao et al.[30] proposed a relationship privacy preservation based on compressive sensing by adopting link randomization technique to achieve the confidentiality of relationship data. To protect against de-anonymization attack, Jianwei Qian et al. [7] construct a comprehensive and realistic model of attacker’s background information with knowledge graphs, for expressing attack process and quantifying privacy disclosure, which provides foundation for generic anynymuzation technique. Daniele Riboni et al.[6] revealed that correlations among sensitive values associated to same individuals in different releases can be easily used to violate user’s privacy by adversaries observing multiple data releases. Thus, they proposed a Jensen–Shannon divergence based defense algorithm to protect user privacy from sequential background knowledge attacks.

Participant coordination is essential in participatory sensing applications, especially when the task’s incentive budget is limited. In real-world applications, rewards should be given to participants, since contributing sensing data costs participants’ bandwidth and battery. Given a certain amount of incentive provided by the task publisher, the system needs to select most participant efficiently under budget constraint to improve the quality of collected data and avoid redundant data collection. Li and Cao have proposed a privacy-preserving incentive distribution mechanism in [22]. Its key idea is to distribute tasks to multiple users and allow users to upload collected data using a pseudonym. However, such method cannot guarantee the QoI achieved. Besides, it may lead to much redundant data, which is not cost effective.

After the sensing data are collected, the server should pay the participants rewards as they have negotiated. To avoid malicious participants who take advantages from uploading wrong data and still get paid, the server should check the correctness of the uploaded data and adjust the rewards of each participant accordingly. Hence, it requires the exact location of the collected data, as well as the identity of the data collector. Wang. et. al. proposed a privacy-preserving reputation update model in [11], which does not rely on a TTP. However, such method still requires that cloaked location of uploaded data, which may lead to inaccuracy in calculating the correctness of the uploaded data.

3. Social participatory sensing system architecture

Based on the most studied scenario in social participatory sensing, we put forward our architecture. As demonstrated in Fig. 1, this architecture comprises of four types of entities: *Application Server (AS), Registration Server (RS), ID-Locator Agent (ILA)* and a set of *Mobile Users (MUs)*. Basically, *AS* is a platform for managing social participatory sensing tasks. *ILA* is an entity which dynamically allocates anonymous identification, i.e., *locator* to *MUs*. *RS* keeps associations between each mobile user's locator ID and its latest IP address. *MUs* are those data contributors who move around over task area, motivated by incentive awards to participate sensing tasks.

Following the basic concepts of traditional models, we still divide the schemes of social participatory sensing into five stages. But being different from previous researches, the key mechanism of privacy-preserving is embodied in the following aspects:

1. In order to avoid disclosing private location information to *AS* for participant selection, mobile users collaborate in Secure Multi-Party Cooperation (MPC) way to report their data sensing ability to *AS* without revealing own trajectory data.
2. In order to reduce the risk that sensing data being associated with its contributor, *inferable privacy protection capability* is taken into account in participant selection stage. More concretely, entropy is adopted to evaluate the inferable privacy level of participant group. Those candidate, whose joining the participant group may bring greater inferable privacy protection capability would take precedence.
3. Participants are selected iteratively in such a way that, each participant joins in the participants set with only communicating with his/her direct parent node and descendant node. We call this structure of participants set as *Participants-Chain (PC)*. PC-based data aggregation and uploading ensures that *AS* is not aware of the contributor of certain piece of sensing data.
4. Moreover, double encrypted incentives distribution along reverse *Participants-Chain* ensures each participant successfully obtain his/her deserved incentives without revealing identity information.
5. Mobile users' ID and IP address are separated in order to avoid single server being able to access mobile user's *visible privacy* (trajectory information).

Two important assumptions need to be clarified for this architecture: (1)motivated by incentive and for his/her own privacy concern, mobile users are willing to join the complicated collaborated task of participant selection; (2) AS, ILA and RS are honest, they would not collude with each other to steal user's privacy. Then, before elaborating the above privacy-preserving mechanisms in three key stages: participants selection, data aggregation and uploading as well as incentive distribution. We will first describe how *ILA* and *AS* work cooperatively for visible privacy preserving. As illustrated in Fig. 2, *ILA* maintains an associative table which records each mobile user's real ID and his/her time-varying locator. *ILA* assigns unique and time-varying locator ID to every mobile user periodically. Mobile user queries his/her locator from *ILA*, and then reports his IP address to *RS* with the latest locator ID and a new randomly generated token. The separation between mobile user's real ID and accessed IP address assures that neither *AS* nor *ILA* has enough knowledge to infer mobile user's trajectory. Fig. 1 also demonstrates the procedure how *AS* acquires a mobile user's IP address. *AS* first asks *ILA* for target mobile user's latest locator ID, then it queries the demanded IP address from *RS* with the token granted by the mobile user. The token will be valid for a single use and would remain in effect for a period between two updates of mobile user's IP address. What matters to set a token is to ensure that *AS* has no right to query any mobile user's IP address without permission. The following sections will elaborate the detail privacy-preserving mechanisms. For convenience, all the notations appearing in this paper are listed in Table 1.

4. Secure Multi-party Cooperation based participant selection

4.1. Preliminary analysis

Suppose *AS* publishes a task to collect data in designate area $\mathscr{L}$. To avoid possible counterfeit sensory data from malicious participants and improve data accuracy, multiple yet most *N* copies of sensing data are required in each region l_i, $(1 \le i \le L)$. To encourage mobile users to join in the sensing task, *AS* provides total *B* incentive to recruit participants. Mobile users $\mathscr{M}$ move around, they see their own trajectory in a near future. $R_i = \{r_i^1, r_i^2, ..., r_i^L\}$, $(1 \le i \le M)$ defines the set of tasks mobile user m_i promises to undertake based on his/her future trajectory. Where,

$$r_i^j = \begin{cases} 1, \text{if } m_i \text{ can collect data in region } l_j \\ 0,\text{otherwise} \end{cases} \tag{1}$$

Mobile user m_i's exception of incentives for R_i is then defined as $REQ_i = \{req_i^l, req_i^2, ..., req_i^L\}$.

The fundamental problem for participant selection is about how to select optimal participants to achieve good QoI under incentive budget constraint, of which the premise is to ensure mobile users' privacy.

4.2. Constrain conditions for social participatory sensing

Participants selection is a multi-constraint conditions based decision problem, *AS* needs to balance various factors to achieve optimal result. Besides the most studied factors: *Data Utility* and *Incentive Budget Constrain. Inferable Privacy Protection Ability* is also taken into account in judging whether a mobile user is suitable for being involved as a participant.

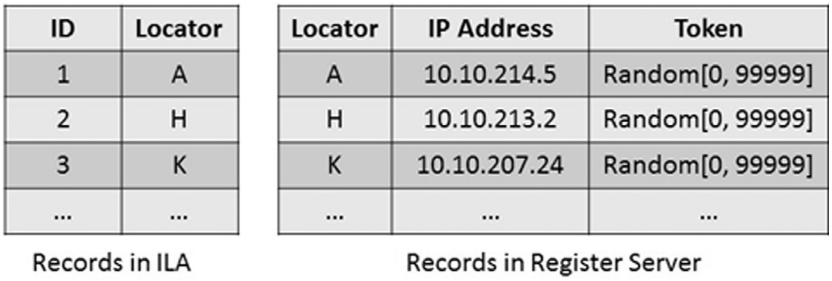
Data Utility. For almost every social participatory sensing task, enough evenly distributed collected data is beneficial for data fusion and knowledge discovering. We borrow the concept of *Data Utility* from Zheng Song, et al.[2] to measure the quality of data set. Let $Q = \{q_1, q_2, ..., q_L\}$ denote the amount of data sensed by participants $\mathscr{P}$ in all regions, where

$$q_j = \sum_{m_i \in \mathscr{P}} r_i^j, (1 \le j \le L) \tag{2}$$

Then, we can give the definition of *data utility*$u(\mathscr{P})$ as:

$$u(\mathscr{P}) = 1 - \frac{\sum_{1 \le j \le L} (N - q_j)^2}{L \times N^2} \tag{3}$$

Data utility reflects both coverage and redundancy of the sensing data set. Obviously, there is a positive correlation between the increment in q_j and the increment in $u(\mathscr{P})$, if condition $q_j \le N$ is satisfied. That is to say, the more sensing data are collected in a region, the greater contribution it would make to the *data utility*. But, if redundancy occurs, namely $q_j > N$, *data utility* decreases. Next, we shall prove that when the total amount of collected data is fixed, *data utility* reaches its maximum value if the amount of data is evenly distributed in each region. For convenience, the restriction that data in each region is evenly distributed is specialized to the situation that the amount of data in each region is equal. So, the problem is simplified as: given $Num = \sum_{m_i \in \mathscr{P}} num_i$,$u(\mathscr{P})$ reaches maximum *iff*$q_1 = q_2 = ,...,=q_L$.



ID	Locator
1	A
2	H
3	K
...	...

Records in ILA

Locator	IP Address	Token
A	10.10.214.5	Random[0, 99999]
H	10.10.213.2	Random[0, 99999]
K	10.10.207.24	Random[0, 99999]
...	...	...

Records in Register Server

Fig. 2. Records in ILA and Register Server.

Table 1
Notations.

Symbols	Descriptions
$\mathscr{M} = \{m_1, m_2, ..., m_M\}$	a set of M mobile users
$\mathscr{L} = \{l_1, l_2, ..., l_L\}$	a set of L locations
$\mathscr{P} = \{p_1, p_2, ..., p_{top}\}$	the set of selected participants
$\mathscr{C} = \{can_1, can_2, ..., can_{M-P}\}$	the set of candidates
Num	the total amount of data all participants collect
$R_i = \{r_i^1, r_i^2, ..., r_i^L\}$	a vector indicates whether m_i can collect data in each location
num_i	the total amount of data m_i collects
$REQ_i = \{req_i^1, req_i^2, ..., req_1^L\}$	m_i's requested incentive in each location
TRI_i	the amount incentive m_i requests
$D_i = \{d_i^1, d_i^2, ..., d_i^L\}$	the data collected by m_i in each location
$Q = \{q_1, q_2, ..., q_L\}$	the amount of data collected by selected participants in each location
N	the maximum amount of data required in a location
B	the total amount of incentive budget for the task
inc_i^l	the incentive allocated to d_i^l
γ_i^l	random number associated to d_i^l
top	the index of the newest selected participant in participants-chain
$u(\mathscr{P})$	the data utility of data collected by $\mathscr{P}$
$H(\mathscr{P})$	the inferable privacy protection ability of $\mathscr{P}$
$\widehat{req_i^l}$	the incentive allocated to d_i^l
$\mathscr{K}_{spub}$ and $\mathscr{K}_{spri}$	the public key and private key generated by AS
$\mathscr{K}_i^l$	the public key generated by m_i for d_i^l
$\mathscr{K}'_{spub}$ and $\mathscr{K}'_{spri}$	the public key and private key generated by AS for incentive distribution

Proof. Actually, this is an optimization problem: maximize $u(\mathscr{P})$, subject to $g(q_1, q_2, ..., q_L) = \sum_{1 \le j \le L} q_j - Num = 0$. Lagrange multipliers is a powerful tool for solving this problem without the need to explicitly solve the conditions and use them to eliminate extra variables. Let $\varphi(q_1, q_2, ..., q_L, \lambda) = u(\mathscr{P}) + \lambda * g(q_1, q_2, ..., q_L)$. Where, λ is Lagrange multiplier. The method of Lagrange multipliers relies on the intuition that $u(\mathscr{P})$ cannot be increasing at a maximum in the direction of any neighboring point where $g(q_1, q_2, ..., q_L) = 0$. If it were, we could walk along $g = 0$ to get higher,meaning that the starting point was not actually the maximum. So, we can take the partial derivatives of each variables of $\varphi(q_1, q_2, ..., q_L, \lambda)$ and make them equal to 0:

$$\begin{cases} \dfrac{d\varphi}{dq_1} = \dfrac{2 \times (N - q_1)}{L \times N^2} + \lambda = 0 \\ \dfrac{d\varphi}{dq_2} = \dfrac{2 \times (N - q_2)}{L \times N^2} + \lambda = 0 \\ \cdots \\ q_1 + q_2 + \ldots + q_L - Num = 0 \end{cases} \tag{4}$$

To solve Eq. (4), we can obtain $q_1 = q_2 = \ldots = q_L = \frac{Num}{L}$, and $\lambda = \frac{2 \times (A - L \times N)}{(L \times N)^2}$. So, it can be concluded that *data utility* reaches its maximum value *iff* the data amount in each region is equal. Based on this investigation, we can further deduce that, the more even the data amount distribution is, the greater the *data utility* can be obtained. Overall, the objective in terms of QoI is to enlarge Eq. (5) by selecting appropriate mobile user.

$$u(\mathscr{P} + m_i),\ m_i \in \mathscr{C} \tag{5}$$

Incentive Budget Constrain Let $REQ_i = \{req_i^1, req_i^2, ..., req_i^L\}$ denote the set of requested incentive of mobile user m_i, and let $TRI_i = \sum_{1 \le j \le L} req_i^l$ denote the total incentive amount m_i requests. Therefore, the incentive budget constrain can be stated as: the total amount of incentive requests in all regions cannot exceed the predefined total budget. We give the formal constraints in Eq. (6):

$$TRI_i + \sum_{m_j \in \mathscr{P}} TRI_j \le B \tag{6}$$

Inferable Privacy Protection Ability. Consider that the collected data is semantic-rich, like photo, video, twitte, or any other User Generated Content (UGC). These data may reveal associations between individuals and sensitive information, for example, occupation, income level or even political tendency. Correlations among sensitive values associated to the a same individual can be easily used to violate user's privacy by adversaries observing multiple pieces of sensory data, even if state-of-art privacy protection techniques are applied. The most common background knowledge attack is based on association rules mining. A pre-condition for this kind of attack is that the adversary can associate multiple pieces of separated information together to target an object person. So, if the *AS* has gathered *Num* pieces of data, then the maximum degree of identity privacy is achieved when it sees all the data as equally probable for providing important clues to associate the deduced identity to a participant with certain background information. Therefore, the degree of inferable privacy level depends on the distribution of the probabilities for the total *Num* pieces of collected data. Following the concept proposed in [24], we adopt entropy, which is a measure of unpredictability in information theory to evaluate the inferable privacy level of the participants group. Greater entropy value makes it harder to predict the association relationship between each sensory data with its contributor. So the objective of our approach is to maximize the entropy so that the collected data is with nearly average probability of being associated with certain participant. Given a series of random variable num_i, $(1 \le i \le top)$, the probability of identifying a collected data belonging to participant P_i is then denoted as $\frac{num_i}{Num}$. So, we defined the entropy H of the participants group as:

$$H(\mathscr{P}) = -\sum_{1 \le i \le top} \left(\frac{num_i}{Num} \cdot \log_2 \frac{num_i}{Num} \right) \tag{7}$$

$H(\mathscr{P})$ measures the uncertainty of verifying a participant from collected data, so a greater $H(\mathscr{P})$ would contribute better inferable privacy protection. □

4.3. Secure Multi-party Cooperation

As discussed in previous section, participants coordination is a complex multi-criterion decision procedure. *AS* makes comprehensive judgement according to mobile user's data utility contribution, incentive request and privacy protection ability. To calculate the incentive request and inferable privacy protection ability, mobile users only report position-independent values to *AS*, which does not contain any sensitive information about location privacy or identity privacy. However, as illustrated in Eq. (5), the calculation of *data utility* needs mobile user's location-related data collection plan R_i as input. Obviously, this require does not conform to mobile user's concern about location privacy. Based on this consideration, we propose a Secure Multi-party Cooperation (SMC) based approach to calculate the *data utility* without disclosing the private inputs to the other parties.

Suppose a group of mobile users have been selected as participants $\mathscr{P}$, the amount of data they can collect in all regions is denoted as $\{q_1, q_2, ..., q_L\}$. $R_i = \{r_i^1, r_i^2, ..., r_i^L\}, (1 \le i \le M)$ indicates whether mobile user m_i can collect data in L regions. So, the *data utility* when m_i is selected is updated as:

$$\begin{aligned} u(\mathscr{P} + m_i) &= 1 - \frac{\sum_{1 \le j \le L} (N - (q_j + r_i^j))^2}{L \times N^2} \\ &= 1 - \frac{(N - (q_1 + r_i^1))^2 + ... + (N - (q_L + r_i^L))^2}{L \times N^2} \\ &= \frac{2 \times N \times \sum_{1 \le j \le L} q_j}{L \times N^2} + \frac{2 \times N \times \sum_{1 \le j \le L} r_i^j}{L \times N^2} \\ &- \frac{\sum_{1 \le j \le L} (q_j)^2 + \sum_{1 \le j \le L} (r_i^j)^2}{L \times N^2} \\ &- \frac{2 \times \sum_{1 \le j \le L} q_j \times r_i^j}{L \times N^2} \end{aligned} \tag{8}$$

We find that except for the last item in Eq. (8), i.e., $\frac{2 \times \sum_{1 \le j \le L} q_j \times r_i^j}{L \times N^2}$, the other items can be calculated by the selected participants $\mathscr{P}$ or the mobile user m_i independently. Actually, the last item is scalar product of two vectors, denoted as $\frac{2 \times \overrightarrow{Q} \cdot \overrightarrow{R_i}}{L \times N^2}$. However, since $\overrightarrow{R_i}$ contains mobile user m_i's trajectory-related information, reporting $\overrightarrow{R_i}$ to *AS* or $\mathscr{P}$ will cause trajectory privacy disclosure. To solve this problem, the Secure Multi-part Coordination mechanism will be elaborated next.

Problem Definition: P_{top}, cluster head of the selected participants group $\mathscr{P}$ holds $\overrightarrow{Q} = [q_1, q_2, ..., q_L]^T$, while mobile user m_i keeps $\overrightarrow{R_i} = [r_i^1, r_i^2, ..., r_i^L]^T$. *AS* is to get the result of $u(\mathscr{P} + m_i)$, while P_{top} and m_i keep their own input private.

The basic concept of Secure Multi-party Cooperation is to give P_{top} a disguised result $\overrightarrow{Q} \cdot \overrightarrow{R_i} + v$ to prevent him/her from knowing the partial result, where v is a random scalar known to m_i only. Later, the effect of v could be effectively eliminated by *AS* after it receives the scalar result of $\overrightarrow{Q} \cdot \overrightarrow{R_i} + v$ and v from the two parties.

To solve this problem, we transform $\overrightarrow{Q}$ and $\overrightarrow{R_i}$ to another vector $\overrightarrow{Q}'$ and $\overrightarrow{R_i}'$, such that the disclosing information about $\overrightarrow{Q}'$ and $\overrightarrow{R_i}'$ would not allow adversary or malicious user to derive the original vectors. A linear transformation can be adopted for this target. Suppose X is an $L \times L$ invertible matrix, let $\overrightarrow{Q}' = \overrightarrow{Q} \cdot X$. Disclosing half of $\overrightarrow{Q}'$ will not allow others to derive the original data. So, based on this foundation, we elaborate the SMC-based approach as below:

1. *AS* sends a random invertible $L \times L$ matrix X to both P_{top} and m_i (for convenience, L is supposed to be even, otherwise,a ”0” element can be padded to $\overrightarrow{Q}$ and $\overrightarrow{R_i}$).
2. P_{top} disguises $\overrightarrow{Q}$ by letting $\overrightarrow{Q}' = \overrightarrow{Q} \cdot X$. Divided equally, $\overrightarrow{Q}'$ is denoted as $\left[\overrightarrow{Q}'_{left} \middle| \overrightarrow{Q}'_{right} \right]^T$. P_{top} sends $\overrightarrow{Q}'_{right}$ to m_i.
3. m_i disguises $\overrightarrow{R_i}$ by letting $\overrightarrow{R_i}' = X^{-1} \cdot \overrightarrow{R_i}^T$. Divided equally, $\overrightarrow{R_i}'$ is then denoted as $\left[\overrightarrow{R_{i_{top}}}'^T \middle| \overrightarrow{R_{i_{bottom}}}'^T \right]$. m_i sends $\overrightarrow{R_{i_{top}}}'^T$ to P_{top}.
4. P_{top} calculates $u = \overrightarrow{Q}'_{left} \cdot \overrightarrow{R_{i_{top}}}'^T$.
5. m_i calculates $v = - \overrightarrow{Q}'_{right} \cdot \overrightarrow{R_{i_{bottom}}}'^T$.

It is easy to deduce that $u = \overrightarrow{Q}' \cdot \overrightarrow{R_i}' + v = \overrightarrow{Q} \cdot \overrightarrow{R_i} + v$. Let $W = [w_1, w_2]^T$, $Z = [z_1, z_2]^T$, where

$$\begin{cases} w_1 = \frac{2 \times N \times \sum_{1 \le j \le L} q_j}{L \times N^2} - \frac{\sum_{1 \le j \le L} (q_j)^2}{L \times N^2} \\ w_2 = \frac{u}{L \times N^2} \\ z_1 = \frac{2 \times N \times \sum_{1 \le j \le L} r_i^j}{L \times N^2} - \frac{\sum_{1 \le j \le L} r_i^j}{L \times N^2} \\ z_2 = \frac{v}{L \times N^2} \end{cases} \tag{9}$$

P_{top} and m_i sends W and Z to *AS* respectively, and then *AS* can obtain the updated *data utility* as:

$$u(\mathscr{P} + m_i) = w_1 + z_1 - 2 \times (w_2 - z_2) \tag{10}$$

4.4. Multi-criteria participants decision

Fig. 3 illustrates the message sequence between all involved entities in participants coordination procedure. First, *AS* will broadcast a $L \times L$ matrix X to all candidates ($can_1, can_2, ..., can_n$) and P_{top}, the cluster head of selected participants $\mathscr{P}$. Second, each candidate and the P_{top} invoke function to disguise their private input collaboratively according to the Secure Multi-party Cooperation method we elaborate before. Third, along with disguised data and the total required incentive, the amount of data he/she expects to collect and a varying token, each candidate reports these data to *AS*. Meanwhile, the cluster head P_{top} also reports the disguised data generated in Secure Multi-party Cooperation process to *AS*. Fourth, *AS* invokes a multi-criteria decision function to select winner from all the candidates. Fifth, *AS* requests winner's locator ID from *ILA*; Sixth, *AS* requests winner's latest IP address with the locator ID and his/her token from *RS*. Finally, the newly selected winner is designated as new cluster head of the selected participants, and its latest IP address is broadcasted to the remaining candidates.

Actually, how to balance *data utility, incentive budget constrain* and *referable privacy protection ability* is indeed a multi-criteria decision problem. To solve this problem, a fuzzy decision based multi-criteria ranking algorithm, i.e., Preference Ranking Organization Method of Enrichment Evaluation (PROMETHEE) [31] is adopted to select the most appropriate candidate as participant.

PROMETHEE method is founded on the analysis of the difference of objects according to constituent criteria. Difference of these values determines the preference one object above another. PROMETHEE method is widely used in multiple criteria decision and ranking area.

Suppose we have a set of alternatives $X = \{x_1, x_2, ..., x_m\}$, and a set of evaluation criteria $C = \{c_1, c_2, ..., c_n\}$. Weights of each criteria is denoted as $W = \{w_1, w_2, ..., w_n\}$, s.t.$\sum_{k=1}^{n} w_k = 1$. Let $a_{i,\,k}$ denote the evaluation of object x_i on criteria c_k, and $g(x_{i,\,k}, x_{j,\,k})$ denote the degree of superiority that x_i surpasses x_j on criteria c_k. So, we can get the aggregated preference value x_i over x_j on all criteria as:

$$\pi(x_i, x_j) = \sum_{k=1}^{n} w_k \times g(a_{i,k}, a_{j,k}) \tag{11}$$

Here, we define $g(x_{i,\,k}, x_{j,\,k})$ as:

$$g(x_{i,k}, x_{j,k}) = \begin{cases} 1 - \exp^{-\frac{(x_{i,k} - x_{j,k})^2}{2\sigma^2}}, (x_{i,k} - x_{j,k}) > 0 \\ 0, (x_{i,k} - x_{j,k}) \le 0 \end{cases} \tag{12}$$

Furthermore, we deduce the positive outranking flow Φ^+ and negative outranking flow Φ^- of candidate x_i as:

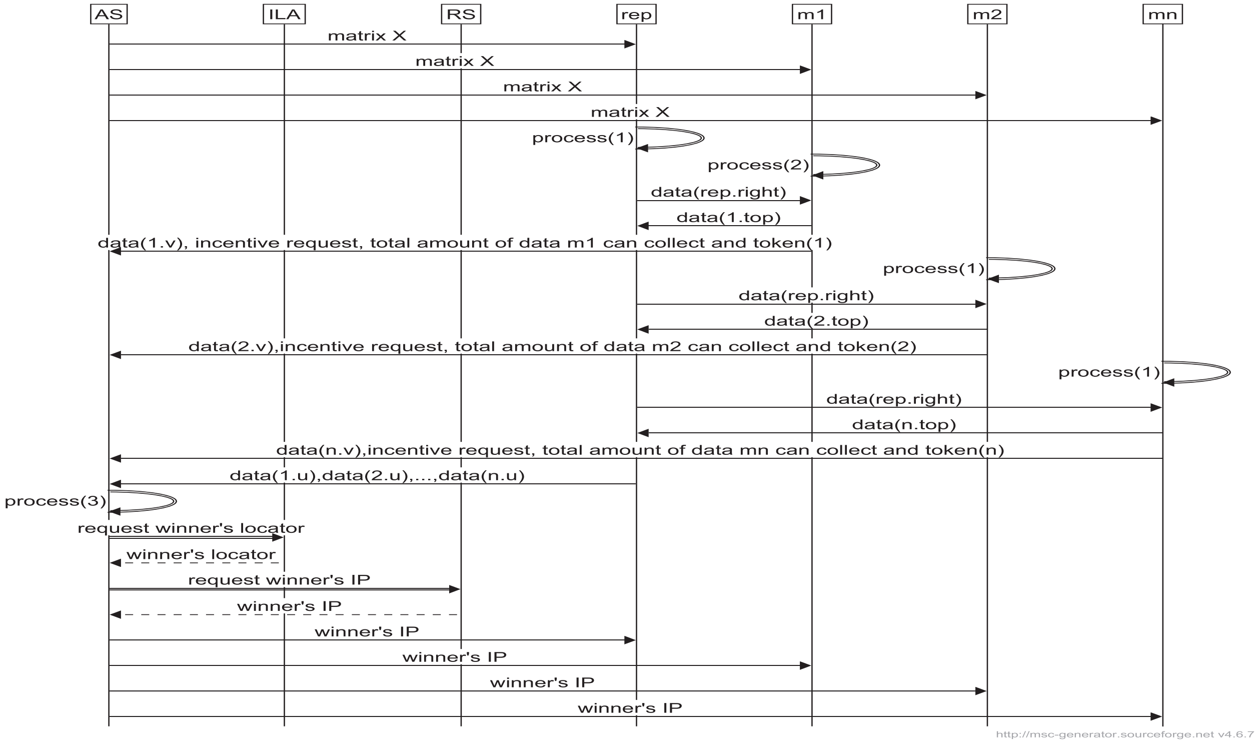

AS
ILA
RS
rep
m1
m2
mn
matrix X
matrix X
matrix X
matrix X
process(1)
process(2)
data(rep.right)
data(1.top)
data(1.v), incentive request, total amount of data m1 can collect and token(1)
process(1)
data(rep.right)
data(2.top)
data(2.v),incentive request, total amount of data m2 can collect and token(2)
process(1)
data(rep.right)
data(n.top)
data(n.v),incentive request, total amount of data mn can collect and token(n)
data(1.u),data(2.u),...,data(n.u)
process(3)
request winner's locator
winner's locator
request winner's IP
winner's IP
winner's IP
winner's IP
winner's IP
winner's IP
http://msc-generator.sourceforge.net v4.6.7

Fig. 3. Participants Coordination Process.

$$\begin{cases} \Phi^{+}(x_i) = \dfrac{\sum_{j=1}^{m} \pi(x_i, x_j)}{m-1} \\ \Phi^{-}(x_i) = \dfrac{\sum_{j=1}^{m} \pi(x_j, x_i)}{m-1} \end{cases} \tag{13}$$

Where, $\Phi^{+}(x_i)$ and $\Phi^{-}(x_j)$ represent how much degree x_i precedes other candidates and how much degree other candidates precedes m_i in overall criteria respectively. Nevertheless, the distribution of weights among different criteria is a key problem affecting the candidate ranking result. In order to alleviate potential subject bias in weight allocation, we adopt entropy weight algorithm to determine each criteria's weight. Entropy is a measure of the unpredictability of information content. Entropy method determine the weight value of each criteria by calculating the entropy. The greater the entropy is, the smaller the corresponding entropy weight is, and then the amount of valuable information it can provide for decision is reduced. Let e_k denote the entropy of the c_k, it can be calculated as:

$$e_k = -\frac{1}{\ln m} \times \sum_{i=1}^{m} f_{ik} \ln(f_{ik}) \tag{14}$$

where $f_{ik} = \frac{a_{i,k}}{\sum_{i=1}^{n} a_{i,k}}$ is the proportion of $a_{i,\,k}$ to the sum of evaluations on all criteria. According to Eq. (12), the entropy weight of the k_{th} criteria c_k is then determined as:

$$w_k = \frac{1 - e_k}{n - \sum_{i=1}^{n} e_i} \tag{15}$$

So, based on the PROMETHEE method, the participants selection procedure is illustrated in Algorithm 1.

As described in Algorithm 1, *AS* first ranks candidate mobile users according to their net outranking flow Φ, and then from top to bottom, it selects the top one as new participant if his/her incentive request is not over the remaining budget. Participant coordination procedure select one participant from candidates in each round. This process continues, until budget is not enough or enough participants have been enrolled. Finally, a *participant-chain*, denoted as $\mathscr{P} = \{p_1, p_2, \ldots, p_{top}\}$, is obtained.

4.5. Computation cost analysis

Suppose at last k participants are selected from m candidates according to n evaluation criteria ($n = 3$). Thus, in each iteration, the computation cost in candidate side is $O(m)$ (corresponding to the process of computing each individual's *data utility gain, incentive request* and *inferable privacy protection ability*). Finally after k iterations, the total computation cost is $O(nk)$, i.e., $O(3k)$. For *AS* side, the computation cost mainly comes from two parts, namely the process of computing each candidate's *data utility gain* and multi-criteria decision based ranking. In each iteration, the computation cost for computing the *data utility gain* is $O(m)$, and the computation cost for multi-criteria decision based ranking is $O(m^2 + nm)$. So after k iterations, the total computation cost is $O((m + m^2 + nm)*k)$, i.e., $O((m + m^2 + 3m)*k)$.

5. Data uploading and incentive distribution

Once participants selection procedure ends, participants carry out data collection task as they have planned. At the end of the task time, participants upload their collected data along with the *participant-chain*. According to the actual received data, *AS* determines each participant's deserved incentive and distribute the incentives along the *participant-chain* reversely. Since collected data is tagged with location, moreover, incentive should be associated to the right participant, effective privacy preserving mechanisms are essential in data uploading and incentive distribution stage.

5.1. IP address updating mechanism

Before elaborating the detail procedure of data uploading and incentive distribution, we would first describe participant's IP address update strategy. As shown in Fig. 4, participant moves along his own trajectory, so his accessible IP address may probably has shifted from the origin. On one hand, *AS* and his/her child node (we call node P_i's

```
Input:
$\{u(\mathcal{P}+m_i)|m_i \in (\mathcal{M}-\mathcal{P})\}$; $\{H(\mathcal{P}+m_i)|m_i \in (\mathcal{M}-\mathcal{P})\}$;
$\{REQ_i|m_i \in (\mathcal{M}-\mathcal{P})\}$;
Output: A new participant,denoted as $P_{top}$.
for each $m_i \in (\mathcal{M}-\mathcal{P})$ do
    compute $\Phi^+(m_i)$ according to Eq. (13)     compute $\Phi^-(m_i)$ according to Eq. (13)     compute $\Phi(m_i) = \Phi^+(m_i) - \Phi^-(m_i)$;
end
sort candidates $(\mathcal{M}-\mathcal{P})$ according to $\Phi$ in descending order to get ordered set:$C = \{can_1, can_2, \ldots, can_{M-P}\}$; while $i \le M - P$
do
    if $B(P) + B(i) \le B$ then
        set $can_i$ as new participant;
        designate $can_i$ as $P_{top}$;
        Break out;
    end
    i=i+1;
end
if $i > M - P$ then
    Return Null;
end
else
    Return $P_{top}$;
end
```

Algorithm 1. Multi-criteria ranking based participants selection.

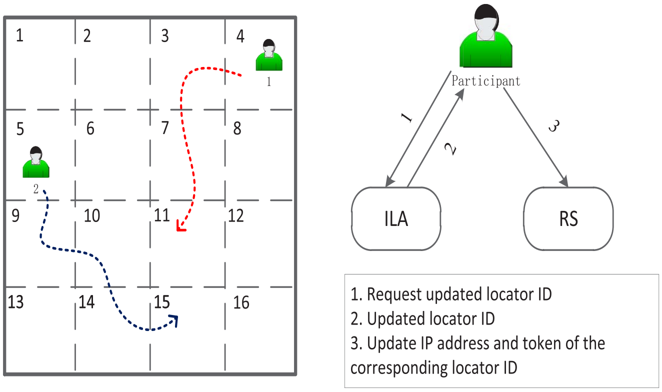

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
1
2
Participant
1
2
3
ILA
RS
1. Request updated locator ID
2. Updated locator ID
3. Update IP address and token of the corresponding locator ID

Fig. 4. IP Address Update Process.

downstream adjacent node P_{i+1} in participant-chain as his/her parent/child node, in reverse, P_i is called P_{i+1}'s parent node) need to exchange information with a participant, on the other hand, participant is not willing to reveal his/her ID-Address association to others for location privacy consideration. So, we deploy *ILA* and *RS* in our model. The basic idea is to isolate mobile user's ID and IP address. At each fixed interval of time, *ILA* will update all mobile users' locator ID with a random function, also *RS* will wipe the IP address records of each mobile user. Then, participant will request *ILA* for the latest locator ID, and then with the locator ID the participant reports his/her new IP address and token to *RS* to register the new accessible IP address.

5.2. Data uploading

Fig. 5demonstrates the message interaction process between related entities. Participant-chain based data uploading and incentive distribution is adopted to hide the association between each sensory data/incentive and its provider/deserver. Two encryption methods are utilized to prevent the collected data and its corresponding incentive from being distorted or misappropriated. Along with the participant-chain, each participant uploads his/her own encrypted data and those received from his/her child node, to his/her parent. Finally, *AS* receives the whole data from the top participant. This process is described in more detail terms as below.

At initial phase, *AS* generates a pair of keys, $\{\mathscr{K}_{spub}, \mathscr{K}_{spri}\}$ for encrypting and decrypting the uploaded data. Then *AS* inquires all participants' latest IP addrss from *RS*, and then it sends them the $\mathscr{K}_{spub}$ and their respective parent's IP address. After that, each participant encrypts each piece of collected data with $\mathscr{K}_{spub}$ according to Asymmetric Cryptographic Algorithm (ACA)[18]. Besides the sensory data d_i^l, the location tag l and the expected incentive req_i^l are also encrypted.

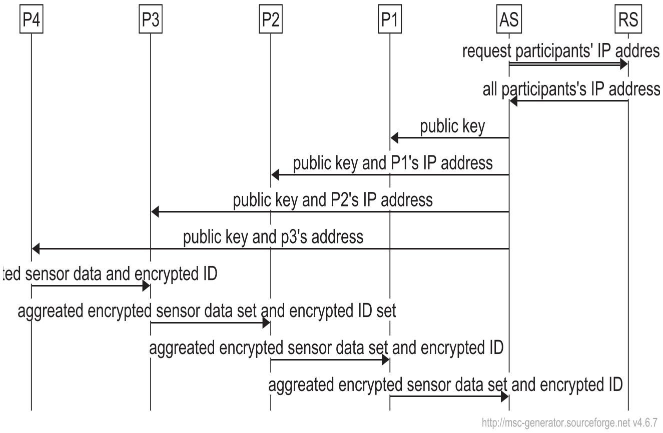

P4
P3
P2
P1
AS
RS
request participants' IP addres
all participants's IP address
public key
public key and P1's IP address
public key and P2's IP address
public key and p3's address
ted sensor data and encrypted ID
aggreated encrypted sensor data set and encrypted ID set
aggreated encrypted sensor data set and encrypted ID
aggreated encrypted sensor data set and encrypted ID
http://msc-generator.sourceforge.net v4.6.7

Fig. 5. Data Uploading Process.

Moreover, a unique random γ_i^l and a public key $\mathscr{K}_i^l$ generated by the participant are encapsulated. So, the encrypted packet is in format of:

$$\{l|d_i^l|req_i^l|\gamma_i^l|\mathscr{K}_i^l\}\mathscr{K}_{spub} \tag{16}$$

In order to help *AS* to validate whether participant has successfully uploaded his/her data and to prevent potential malicious activities from other users, each participant is requested to upload an extra encrypted message indicating his/her ID. Meanwhile, a token is also appended in this packet for *AS* acquiring his/her latest IP address. So, the extra encrypted packet is in format of:

$$\{ID_i|token_i\}\mathscr{K}_{spub} \tag{17}$$

It's worth nothing that all participants mix their own encrypted sensory packets and ID packet randomly with those received form his/her child before sending to parent. So, *AS* or upstream nodes cannot deduce the association between any participant and his/her data. After receiving the encrypted data from the first selected participant, *AS* decrypts each encrypted packet with $\mathscr{K}_{spri}$. All sensory data collected as same location are aggregated for quality evaluation. Here, the majority vote method [32] is an option for detecting the low quality data. Meanwhile, by checking the participants' ID, *AS* can deduce whether any malicious participant has discarded data coming from downstream nodes.

5.3. Incentive distribution

Suppose the incentive for sensory data d_i^l is determined as $\widehat{req_i^l}$. To ensure each participant obtain all his/her deserved incentive without leaking any sensitive information, we adopt Advanced Encryption Standard (AES) [33] to encapsulate each packet. Before distributing incentives to participants, *AS* generates a pair of keys $\{\mathscr{K}'_{spriv}, \mathscr{K}'_{spub}\}$ for encrypting and decrypting the packets, meanwhile it quires all participants' latest address from *RS*. Then *AS* encapsulates each incentive in format of :

$$\gamma_i^l \left| \left\{ \left\{ \widehat{req_i^l} \right\} \mathscr{K}'_{spri} \right\} \mathscr{K}_i^l \right. \tag{18}$$

Here, γ_i^l is appended for participant identifying his/her deserved incentives from a set of encrypted packets. The incentive packets are double encrypted. The outer layer is encrypted with participant's own encryption key $\mathscr{K}_i^l$, which ensures incentives cannot be seized even if it is intercepted by malicious participants. While the inner layer is encrypted with *AS*'s private key $\mathscr{K}'_{spub}$, which is designed to ensure that each participant can successfully obtain all his/her incentives before *AS* publicizes $\mathscr{K}'_{spub}$. Then *AS* sends all incentive packets to the bottom participant and informs all participants on participants-chain his/her child's latest IP address. Each participant picks out his/her own packets according to the appended random number and then sends the other packets to his child. This procedure continues until the last participant received all his packets. It is important to note that, if any participant finds the amount of incentive packets is less than it was supposed to be, he/she will inform *AS* an alert. Only the top node on participants-chain obtains all his/her deserved incentive packets, he/she inform *AS* of a message with his/her identification. Then *AS* publicizes $\mathscr{K}'_{spub}$ to all participants. With this encryption key, all participants decrypt the inner layer of incentive packet to get his/her deserved incentive. Fig. 6 demonstrates the detail message interactions between related entities.

6. Privacy analysis and experiment evaluation

6.1. Privacy analysis

In this section, we would analyze how the proposed scheme can provide privacy-preserving in the key procedures of:1) participants selection, 2) data uploading and incentive distribution.

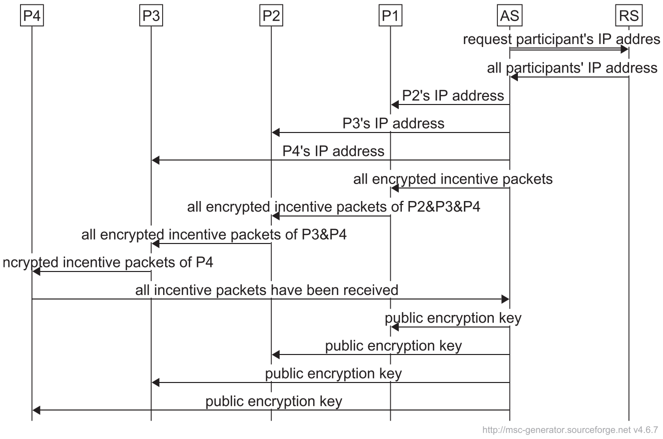

P4
P3
P2
P1
AS
RS
request participant's IP addres
all participants' IP address
P2's IP address
P3's IP address
P4's IP address
all encrypted incentive packets
all encrypted incentive packets of P2&P3&P4
all encrypted incentive packets of P3&P4
ncrypted incentive packets of P4
all incentive packets have been received
public encryption key
public encryption key
public encryption key
public encryption key
http://msc-generator.sourceforge.net v4.6.7

Fig. 6. Incentive Distribution Process.

6.1.1. *Privacy analysis of participant selection*

To analyze the security of the participant selection procedure, we need to prove that neither party,i.e., mobile user m_i and the cluster head P_{top}, has chance to know each other's information. For the selected participants side, P_{top} only knows $L/2$ amount of items in vector $\overrightarrow{R_i}'$. Actually, $X^{-1}\cdot\overrightarrow{R_i}^T$ is a linear transformation to $\overrightarrow{R_i}$, while it is impossible for P_{top} to solve L-dimension linear equations with only $L/2$ equations being obtained. Which is to say, P_{top} could never know the actual values of m_i's private input. For the mobile user m_i, it's the same the other way round. Meanwhile, application server *AS* knows nothing about both parties' detail trajectory-related information. Thus, Secure Multi-party Cooperation based QoI calculation procedure is proved to be effective in privacy-preserving.

6.1.2. *Privacy and security analysis of data uploading and incentive distribution*

Next, we would analyze how the designed participants-chain based data uploading and incentive distribution mechanism ensure participants' privacy.

1) Information interactions between parent and child are anonymized. Either communication party only knows the IP address of the corresponding node, so each participant cannot know other's real identity. Meanwhile, illegal user cannot join in the participant chain, since no one except parent/child is aware of each other's IP address.
2) Each field encapsulated in the uploaded packet does not contain any information specific to a participant oneself, so it is impossible for *AS* to group sensory data by providers, in turn leading to trajectory privacy leakage.
3) Since the association between participants' ID and sensory data is cut-off, *AS* is unable to deduce participants' inferable privacy.
4) Examination on participants' ID provides *AS* information to validate the legality and integrity of received sensory data. Maliciously discarding data received from child node can be detected.
5) Two-layer encryption ensures all participants can get his/her deserved incentives. Malicious interception of other's incentive packet would cause all incentive packets invalidated due to lack of public key.

6.2. *Experiment set up*

We evaluate the proposed scheme by simulations on the Microsoft Research Asia Geolife dataset [34], where real mobility trace of mobile users are used to represent all candidates in social sensing scenario. We adopt the following procedures to set up our simulation:

1) As all traces were spread in different parts of Beijing, a specific rectangular region where the traces mostly appear is needed. We store all trajectories in a geographical MySQL database and find a 200m × 200m region as shown in Fig. 7. We use this region as the simulation area for the considered data collection application.
2) The entire region is divided into 4 × 10 areas of 50m × 50m, i.e., $L = 40$.
3) All 612 trajectories in the considered region are taken as potential (candidate) participants, i.e., $M = 612$. Since these traces are recorded at different times, in our simulation we simply neglect their time index and overlay them into the same time period. Fig. 7 (b) shows the trajectories of all 612 users.
4) The incentive request for each piece is a randomly generated integer ranging between $[min_i, max_i]$, $(1 \le i \le L)$.

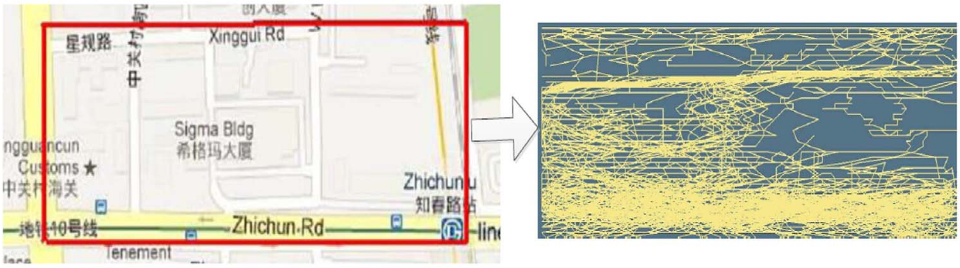
Several experiments are conducted to validate the performance of the scheme we proposed in this paper (referred as UPB-S). Specifically, we compare UPB-S with the following participants coordination scheme:

1) *R-S*: Randomly select a participant in each round until total budget is exhausted;
2) *U-S*: Greedily select a participant who contributes most to the data utility in each round, until total budget is exhausted;
3) *P-S*: Greedily select a participant who brings highest privacy metric in each round, until total budget is exhausted;
4) *B-S*: Greedily select a participant who requests fewest incentive in each round, until total budget is exhausted;

6.3. *Performance*

First, we investigate the performance of different schemes in terms of the index of data utility. Fig. 8 demonstrates the variation trend of data utility with increasing incentive budget. The maximum data volume required in each region is set to 10,20 and 30 respectively.

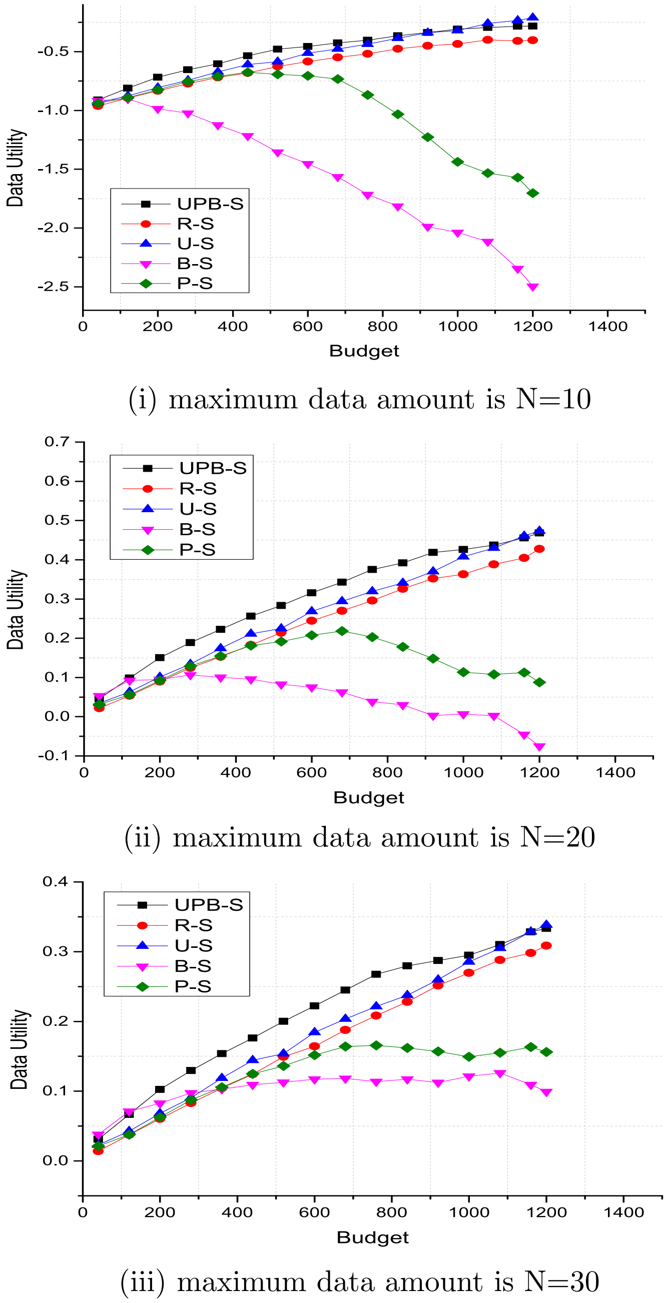
We find that as the incentive budget increases, data utility increases gradually at the initial stage. The best performing scheme is *UPB-S*, followed by *U-S*. *B-S* performs worst in all five schemes. Since *data utility* measures the quality of collected data set from two aspects: the sheer volume and the uniformity of data in all regions. Increasing



(a) Simulation Region

(b) User Trajectories

Fig. 7. Simulation Region with 612 User Trajectories.

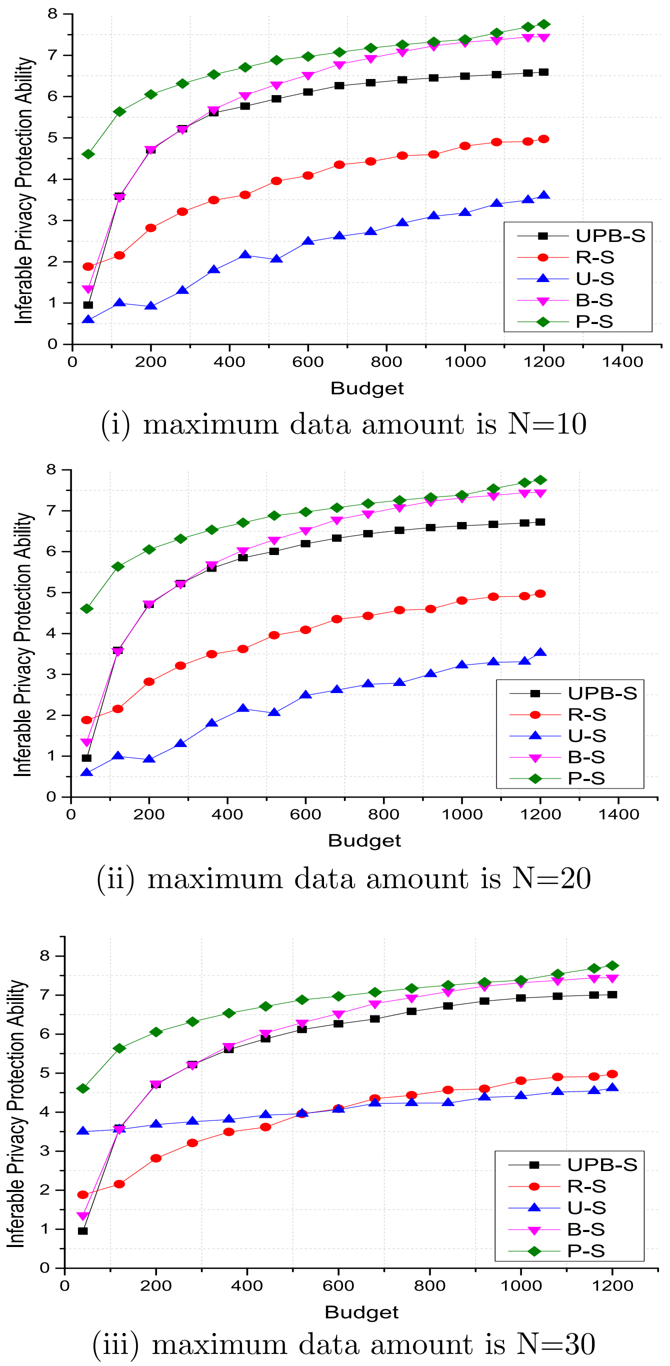


(i) maximum data amount is N=10

(ii) maximum data amount is N=20

(iii) maximum data amount is N=30

Fig. 8. Data Utility with Different Budget.



(i) maximum data amount is N=10

(ii) maximum data amount is N=20

(iii) maximum data amount is N=30

Fig. 9. Privacy Index with Different Budget.

budget motivates more candidates to contribute their data. However, increasing incentive budget also leads to data redundancy in certain regions. Once this situation occurs in a large number of regions, *data utility* declines and finally becomes negative. *UPB-S* performs best among all schemes, since it balances three factors. However, it is interesting that *U-S* just achieves sub-optimal result. The underlying reason is that *U-S* over-pursues individual's contribution which may make the amount of data distribute unevenly. *data utility* goes downhill with incentive increases with *B-S* scheme. The reason is obvious, *B-S* select participant with the goal of saving costs, which makes the supply of data exceed the required amount within budget constrain become possible. *P-S* just pursues the uniformity of data among participants, but ignores the coverage and redundancy, so with the increasing of budget, *data utility* decreases due to data redundancy occurring in a number of regions. Quite unexpectedly, *R-S*'s performance is just after *U-S*. A reasonable explanation is that the random selection policy makes data distribute evenly, so with the increasing of budget, the *data utility* arises.

In the second experiment, we aim to compare different participant coordination schemes in terms of privacy-preserving ability. As shown in Fig. 9, with the increasing of budget, the privacy increases accordingly. That is because the more participants are selected, the less one of the them could be identified due to his/her uniqueness. So, five schemes have risen in varying degrees of privacy level with the increasing of budget. *P-S* always provides optimal referable privacy preserving ability. *B-S* achieves sub-optimal result, since it selects those candidates with the minimum incentive expectation. Actually, these candidates collect very few data, so more participants can be involved with same budget. Thus, *B-S* can achieve better privacy preserving level than others. Although *UPB-S* achieves a result close to *B-S*. Compared with *P-S, B-S* and *UPB-S*, the other two schemes do not obtain good privacy-preserving level. Moreover, *U-S* performs even worse than *R-S* at initial stage. A reasonable interpretation can be that, since *U-S* selects participants who can contribute more data to enlarge the *data utility*, so the evenness of data would obviously decline.

From the above experiments, we can see *UPB-S* obtains optimal data utility while achieves preferable inferable privacy protection level.

7. Conclusion

In this paper, our scheme is proposed to address the privacy problems of participatory sensing, in terms of both visible privacy and inferable privacy, from an novel perspective based on Secure Multiparty Cooperation and fuzzy multi-criteria ranking. It also achieves a quasi-optimal quality requirement of sensing task under the budget constraint of incentives. Analysis of privacy and security indicates that our scheme can well fulfill the privacy-preserving and further detect the malicious participants to propel the task-executing among participants. According to the real trajectories of ordinary citizens in Beijing, the good performance of our scheme is also validated empirically with the comparison of different schemes. It leads to not only an optimal data utility ratio to achieve the quality requirement of sensing task under the budget constraint of incentives, but also a very good inferable privacy level to avoid exposing any individual from the masses of participants. In the future, we plan to further consider the reputation of participants for privacy-preserving participant coordination. Extensively, we will consider not only the privacy protection issue but also the security issues in such systems.

Acknowledgments

We thank Yunjuan Yang and Yining Teng for their hard work and fruitful discussion. This work was supported in part by the National Natural Science Foundation of China under Grant 61602051, 61402045, 61300220 and 61772194, by the Fundamental Research Funds for the Central Universities under Grant 2017RC11, by Open Foundation of State Key Laboratory of Networking and Switching Technology (Beijing University of Posts and Telecommunications) under Grant SKLNST20170202, by the the Scientific Research Fund of Hunan ProvincialEducation Department under Grant No. 16B089, and by the National High Technology Research and Development Program under Grant 2015AA016101.

References

[1] D. Christin, A. Reinhardt, S.S. Kanhere, M. Hollick, A survey on privacy in mobile participatory sensing applications, J. Syst. Software 84(11) (2011) 1928–1946.
[2] Z. Song, C.H. Liu, J. Wu, J. Ma, W. Wang, QoI-aware multitask-oriented dynamic participant selection with budget constraints, IEEE Trans. Veh. Technol. 63(9) (2014) 4618–4632.
[3] F. Hao, M. Jiao, G. Min, L. Yang, A trajectory-based recruitment strategy of social sensors for participatory sensing, Commun. Mag. IEEE 52 (12) (2014) 41–47.
[4] C. Bisdikian, L.M. Kaplan, M.B. Srivastava, D.J. Thornley, D. Verma, R.I. Young, Building principles for a quality of information specification for sensor information, International Conference on Information Fusion, (2009), pp. 1370–1377.
[5] M. Zhang, P. Yang, C. Tian, S. Tang, X. Gao, B. Wang, F. Xiao, Quality-aware sensing coverage in budget-constrained mobile crowdsensing networks, IEEE Trans. Veh. Technol. 65 (9) (2016) 7698–7707.
[6] D. Riboni, L. Pareschi, C. Bettini, Js-reduce: defending your data from sequential background knowledge attacks, IEEE Trans. Dependable Secure Comput. 9 (3) (2012) 387–400.
[7] J. Qian, X.Y. Li, C. Zhang, L. Chen, T. Jung, J. Han, Social network de-anonymization and privacy inference with knowledge graph model, IEEE Trans. Dependable Secure Comput. PP (99) (2017). 1–1
[8] K.L. Huang, S.S. Kanhere, W. Hu, Preserving privacy in participatory sensing systems, Comput. Commun. 33(11) (2010) 1266–1280.
[9] H. To, G. Ghinita, C. Shahabi, A framework for protecting worker location privacy in spatial crowdsourcing, Proc. VLDB Endowment 7(10) (2014) 919–930.
[10] M. Shin, C. Cornelius, D. Peebles, A. Kapadia, D. Kotz, N. Triandopoulos, Anonysense: a system for anonymous opportunistic sensing, Pervasive Mob. Comput. 7(1) (2011) 16–30.
[11] X. Wang, W. Cheng, P. Mohapatra, T. Abdelzaher, Artsense: anonymous reputation and trust in participatory sensing, Proceedings of the 32th IEEE International Conference on Computer Communications (InfoCom), (2013), pp. 2517–2525.
[12] T. Ma, Y. Zhang, J. Cao, J. Shen, M. Tang, Y. Tian, A. Al-Dhelaan, M. Al-Rodhaan, Kdvem: a k-degree anonymity with vertex and edge modification algorithm, Computing 97 (12) (2015) 1165–1184.
[13] H. Rong, T. Ma, M. Tang, J. Cao, A novel subgraph k^+-isomorphism method in social network based on graph similarity detection, Soft. Comput. (2017) 1–19.
[14] K. Vu, R. Zheng, J. Gao, Efficient algorithms for k-anonymous location privacy in participatory sensing, INFOCOM, 2012 Proceedings IEEE, (2012), pp. 2399–2407.
[15] Y. Tian, W. Wang, J. Wu, Q. Kou, Z. Song, E.C.-H. Ngai, Privacy-preserving social tie discovery based on cloaked human trajectories, IEEE Trans. Veh. Technol. 66 (2) (2017) 1619–1630.
[16] Z. Fu, F. Huang, K. Ren, J. Weng, C. Wang, Privacy-preserving smart semantic search based on conceptual graphs over encrypted outsourced data, IEEE Trans. Inf. Forensics Secur. 12 (8) (2017) 1874–1884.
[17] Z. Xia, X. Wang, X. Sun, Q. Wang, A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data, IEEE Trans. Parallel Distrib. Syst. 27 (2) (2016) 340–352.
[18] J. Krumm, A survey of computational location privacy, Pers. Ubiquitous Comput. 13(6) (2009) 391–399.
[19] R.H. Rahman, R. Islam, I. Altas, A.N. Mahmood, J. Abawajy, Withdrawn: an adaptive path hiding technique for participatory sensing networks, Future Gener. Comput. Syst. (2014).
[20] S. Gao, J. Ma, W. Shi, G. Zhan, C. Sun, Trpf: a trajectory privacy-preserving framework for participatory sensing, IEEE Trans. Inf. Forensics Secur. 8(6) (2013) 874–887.
[21] E. De Cristofaro, C. Soriente, Extended capabilities for a privacy-enhanced participatory sensing infrastructure (pepsi), IEEE Trans. Inf. Forensics Secur. 8(12) (2013) 2021–2033.
[22] Q. Li, G. Cao, Providing privacy-aware incentives for mobile sensing, Proceedings of the 11th IEEE International Conference on Pervasive Computing and Communications (PerCom), (2013), pp. 76–84.
[23] J. Shao, R. Lu, X. Lin, Fine: a fine-grained privacy-preserving location-based service framework for mobile devices, Proceedings of the 33th IEEE International Conference on Computer Communications (InfoCom), (2014), pp. 244–252.
[24] I. Boutsis, V. Kalogeraki, Privacy preservation for participatory sensing data, Proceedings of the 11th IEEE International Conference on Pervasive Computing and Communications (PerCom), (2013), pp. 103–113.
[25] R.K. Ganti, N. Pham, Y.-E. Tsai, T.F. Abdelzaher, Poolview: stream privacy for grassroots participatory sensing, Proceedings of the 6th ACM conference on Embedded Network Sensor Systems, (2008), pp. 281–294.
[26] M. Mun, S. Reddy, K. Shilton, N. Yau, J. Burke, D. Estrin, M. Hansen, E. Howard, R. West, P. Boda, Peir, the personal environmental impact report, as a platform for participatory sensing systems research, Proceedings of the 7th ACM conference on Mobile Systems, Applications, and Services, (2009), pp. 55–68.
[27] R. Heatherly, M. Kantarcioglu, B. Thuraisingham, Preventing private information inference attacks on social networks, IEEE Trans. Knowl. Data Eng. 25 (8) (2013) 1849–1862.
[28] T. Jung, X.Y. Li, W. Huang, J. Qian, L. Chen, J. Han, J. Hou, C. Su, Accounttrade: accountable protocols for big data trading against dishonest consumers, IEEE INFOCOM, (2017).
[29] J. Qian, F. Qiu, F. Wu, R. Na, G. Chen, S. Tang, Privacy-preserving selective aggregation of online user behavior data, IEEE Trans. Comput. 66 (2) (2017) 326–338.
[30] L. Yao, D. Liu, X. Wang, G. Wu, Preserving the relationship privacy of the published social-network data based on compressive sensing, IEEE/ACM International Symposium on Quality of Service, (2017).
[31] X. Yu, Z. Xu, Y. Ma, Prioritized multi-criteria decision making based on the idea of promethee, Procedia Comput. Sci. 17 (2013) 449–456.
[32] L.S. Penrose, The elementary statistics of majority voting, J. R. Stat. Soc. (1946) 53–57.
[33] F.P. Miller, A.F. Vandome, J. McBrewster, Advanced encryption standard(2009).
[34] Y. Zheng, X. Xie, W.-Y. Ma, Geolife: a collaborative social networking service among user, location and trajectory, IEEE Data Eng. Bull. 33(2) (2010) 32–39.
[35] X. Wang, M.P. Wellman, Spoofing the limit order book: An Agent-based Model, Conference on Autonomous Agents and Multiagent Systems, International Foundation for Autonomous Agents and Multiagent Systems, (2017), pp. 651–659.