Privacy-preserving QoI-aware participant coordination for mobile crowdsourcing

Bo Zhang\textsuperscript{a}, Chi Harold Liu\textsuperscript{b,\ast}, Jianyu Lu\textsuperscript{c}, Zheng Song\textsuperscript{a}, Ziyu Ren\textsuperscript{d}, Jian Ma\textsuperscript{a}, Wendong Wang\textsuperscript{a}

\textsuperscript{a} State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, China
\textsuperscript{b} School of Software, Beijing Institute of Technology, Beijing 100081, China
\textsuperscript{c} School of Computer Science and Engineering, Huazhong University of Science and Technology, Wuhan, China
\textsuperscript{d} School of Information Science and Technology, Tsinghua University, China

Abstract

Mobile crowdsourcing systems are important sources of information for the Internet of Things (IoT) such as gathering location related sensing data for various applications by employing ordinary citizens to participate in data collection. In order to improve the Quality of Information (QoI) of the collected data, the system server needs to coordinate participants with different data collection capabilities and various incentive requirements. However, existing participant coordination methods require the participants to reveal their trajectories to the system server which causes privacy leakage. But, with the improvement of ordinary citizens’ consciousness to protect their rights, the risk of privacy leakage may reduce their enthusiasm for data collection. In this paper, we propose a participant coordination framework, which allows the system server to provide optimal QoI for sensing tasks without knowing the trajectories of participants. The participants work cooperatively to coordinate their sensing tasks instead of relying on the traditional centralized server. A cooperative data aggregation, an incentive distribution method, and a punishment mechanism are further proposed to both protect participant privacy and ensure the QoI of the collected data. 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 and related work

With the development of smart devices, e.g., smartphones, iPad, and wearable devices, the Internet of Things (IoT) becomes more and more popular and has become the conceptual core of smart home, smart city, etc. As the important sources of information for IoT, mobile crowdsourcing systems also become popular in recent years [4].

In mobile crowdsourcing systems, those people who act as task publishers can post their tasks to online communities or any other channels as they want, and the participants will collect sensory data to complete the sensing tasks. Mobile crowdsourcing paradigm can be applied to a wide range of activities, and even applicable to specific requests, such as crowd-funding, crowd-testing, etc. Cardone et al. [5] initialized a project, called “Partici- pAct Living Lab testbed”, as an ongoing experiment at the University of Bologna involving 300 students for one year in crowd sensing campaigns. It can passively access

Please cite this article as: B. Zhang et al., Privacy-preserving QoI-aware aware participant coordination for mobile crowdsourcing, Computer Networks (2016), http://dx.doi.org/10.1016/j.comnet.2015.12.022
smartphone sensors and also require active user collaborations. Pankratius et al. [6] discussed a crowdsourcing application for space weather monitoring, called “the Mahali project”. Mahali used GPS signals that penetrate the ionosphere for science rather than positioning. A large number of ground-based sensors will be able to feed data through mobile devices into a cloud-based processing environment [7,8], enabling a tomographic analysis of the global ionosphere at the unprecedented resolution and coverage. The authors in [9] studied the platforms for multiple sensing tasks, and proposed a procedural programming language for collecting multiple types of sensor data from a large number of mobile phones [10]. MEDUSA [11] synthesized participatory sensing and crowdsourcing, and put forward a runtime system for multiple sensing tasks with the following stages: task submission, worker selection, and monetary incentive management. In [12], the authors developed a selection framework to enable organizers to identify well-suited participants for data collection, based on both geographic and temporal availability as well as participation habits, source location privacy scheme (SLP). SLP is even more important in military, homeland security, and law enforcement, in addition to many civilian applications [2–4].

Our research is motivated by the application scenario shown in Fig. 1, which is also derived from the smartphone-based environmental monitoring system described in [13]. It shows a population of mobile users subscribing to a registration server as potential participants and a number of task publishers who publish sensing tasks to the application server with its task budget and requirements, which can be characterized by a number of attributes including the sensing locations, sensing time period, and required amount of data at each location and incentive budgets. After data collection, the participants upload sensory data tagged with locations to the application server.

Fig. 1. Considered crowdsourcing sensing scenario.

To achieve better Quality of Information (QoI) for sensing tasks, the application server needs to coordinate appropriate participants for data collection. The approach “DPS” proposed in [14] assumes that the application server knows the exact locations of all potential participants as a prior, and selects a portion of participants to collect more uniformly distributed sensing data within the incentive budget constraint and avoid redundant readings. 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 accordingly. However, the participants’ personal trajectories are often considered as private information. Exposing the trajectory information to the application server may prevent many people from joining participatory sensing [15].

There has been two kinds of approaches to resolve the conflict between the server’s requirements of knowing the participants’ locations and the participants’ requirements of keeping their location privacy. Some assume that there is a trustful third party (TTP) server which is responsible for connecting locations and identifications [16,17]. However, this approach relies too much on the TTP as argued by recent approaches [18,19]. Since the TTP knows too much sensitive information of participants, thus some participants may continue to worry about their privacy and it may become the single target of attacks easily. Therefore, most recent solutions are based on the second approach. The main idea of it is to tradeoff the location accuracy of uploaded data for location privacy. 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 misjudgment of participants’ uploaded data.

Privacy preservation is an important research issue in mobile crowdsourcing, especially for a portion of participants who do not want to share their location data with others [20,21]. The authors in [22] categorized...
privacy into two categories: data privacy and context privacy. The authors in [23] reviewed existing privacy preservation methods and categorized them into six groups: (1) pseudonymity, (2) cloaking, (3) perturbation, (4) sensitive location hiding, (5) exchange-based, and (6) encryption-based methods. The pseudonymity method [24] is used to cut off connections between one’s uploaded data (including trajectory) and his/her identity, and it relies heavily on a TTP. The cloaking-based method [25,26] has been widely adopted recently, and its main idea is to hide the accurate location of data in a cloaked region. The perturbation method [27] and the sensitive location hiding method [28] are designed for distorting the sensor readings (or trajectories) by adding artificial noise. The exchange-based method [29] has been used for data aggregation. Mobile devices exchanges their sensing data labeled with location information before uploading them to the server, so that the server cannot tell the exact trajectory of each participant. This method does not suffer from single point of failure, but the server may fail to evaluate the contribution of each participant. The encryption-based method [30] is frequently used in access control and location privacy protection of the data user. It also relies heavily on a registration center (or a TTP). Such method always leads to loss of data accuracy. The authors in [31] presented a systematic participatory-sensing-based high quality map generation scheme to meet the privacy demand of individual users. The authors in [32] proposed a framework that comprised context-aware vehicular security mechanisms, and collects data from sensors. The information is protected through encryption, authentication, and access control.

Participant coordination is essential in mobile crowdsourcing systems, especially when there is limited incentive budget for the sensing task. In sensing tasks, incentive should be given to participants as their rewards, 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 efficient participant 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 [18]. 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. In [33], assuming that the incentive requests from participants and the defined utility of sensory data on all locations are known, the authors proposed to select a subset of people with maximum sensory data utility. In [34], the authors further improved [33] by studying how to calculate sensory data utility on a certain location. Both research efforts concentrate on selecting participants to maximize the difference between value and price of sensory data. According to [35], the accuracy of sensing results is highly related to the number of collected samples, and thus the application server tends to recruit as many participants as possible to improve the QoI.

After the sensing data are collected, the application server should provide the incentive to the participants as their rewards for data contribution. To avoid malicious participants who take advantages from uploading wrong data and still get paid, the application 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 [26], 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.

In this paper, we propose a privacy-preserving participant coordination mechanism to solve the above problems. The basic idea behind such procedure is to replace the traditional centralized participant coordination phase by cooperation among participants. On the application server side, it selects participant iteratively according to their uploaded related information instead of detailed trajectories. On the participant side, the selected participant of each round passes down not yet satisfied task requirements to other participants. In addition, a data aggregation mechanism, an incentive and punishment mechanism are proposed. The proposed procedures can achieve optimal QoI for sensing tasks, guarantee the robustness of data collection, and preserve the privacy of participants. The major contribution of this paper is four-fold:

- To the best of our knowledge, we are the first to propose a privacy-preserving participant coordination mechanism, which can achieve optimal QoI for sensing tasks without letting the application server know the trajectories of participants.
- We propose a participant selection strategy called “BPS”. The participants are selected by a Borda Ranking based algorithm that explicitly considers their expected capability to fulfill dynamic task requirements, their incentive requirements, and their impact on the privacy-protection level.
- We introduce a distributed mechanism to support data aggregation and incentive distribution that works with the proposed privacy-preserving participant coordination method.
- We analyze the privacy of the proposed methods. Through simulations using real trajectories, we demonstrate the efficiency and robustness of our proposed methods.

In contrast to the existing solutions, our proposed method can guarantee the privacy of participants during the whole procedure of data collection, without wasting incentive or lessening the accuracy of location of uploaded data.

The rest of the paper is organized as follows: Section 2 establishes a formal model of our system. Section 3 describes in details the proposed participant coordination mechanism, and the supporting mechanism for data aggregation, incentive distribution, and punishment mechanism. Section 4 further analyzes the privacy of the proposed mechanisms. Section 5 evaluates the performance of the proposed scheme by extensive simulations using real mobility traces. Finally, Section 6 concludes the paper.
2. System model

2.1. System architecture

This section presents a formal model for describing our participant selection system. We consider a scenario of a certain crowdsourcing application as shown in Fig. 1. The application server consists of an application server, a registration server, and some other servers for data fusion, etc. We consider that a specific task with incentive budget of B is being completed by such a system. In the sensing region, a set of M candidate participants are moving during the time period and collecting data for a specific sensing task, denoted as \( M = m = 1, 2, \ldots, M \). The task requires sampling at multiple locations \( L = l = 1, 2, \ldots, L \) to be taken within a limited time slot \( T \). It is worth noting that, multiple copies of data are required from different participants in each location and time slot in order to avoid error readings from malicious participants and improve data accuracy [36], yet at most \( N \) samples are needed in a specific location and a time slot to avoid too redundant data.

The process of our proposed strategy consists of six parts:

Step 1. The participants register and upload their required incentive to the application server.

Step 2. The task publishers post sensing tasks with requirements and incentive budgets through the application server.

Step 3. The participants accept the sensing tasks and submit their QoI satisfaction ratio increment to the application server, then the application server executes the participant coordination procedure.

Step 4. The selected participants collect and upload the required sensory data to the application server.

Step 5. The application server performs the data fusion process and evaluates the quality of uploaded sensory data.

Step 6. According to the amount and quality of the uploaded data and required incentive, the application server decide the rewards to the participants. And if a participant is judged to be negligent, he/she will be punished by the application server.

Data fusion and data quality evaluation can be provided by applying majority voting method [36] or other approaches. It is not the focus of our work. Instead, we will focus on how coordination can be done while keeping the future locations of participants private in step 3. Also, we will show how the sensing data are collected with accurate location tags without revealing the data collector’s trajectories in step 4; and how the participants are rewarded or punished according to their uploaded data without revealing each piece of data is collected by whom in steps 5 and 6.

In this paper, we assume that all mobile nodes can be accessed through their IP addresses in 3G networks or WiFi networks. In order to relieve the communication load of the application server, we further deploy a registration server which can (1) maintain the real time IP addresses of all devices; (2) feedback the new IP address of a device to inquiry when the IP address has changed.

2.2. Problem formulation for participant coordination

2.2.1. Data utility

Let \( u_m \) denote the data collection utility of a participant \( m \in M \).

\[
u_m = \{u_m^l|\forall l \in L, \forall t \in T\}, \quad \forall m \in M,\]

(1)

where \( u_m^l \) is how many data copies that participant \( m \) can collect in location \( l \) and time slot \( t \).

Then we could calculate the utility of a group of participants. Let \( X \subset M \) denote the set of selected participants. The data collection utility \( U(X) \) can be denoted as:

\[
U(X) = \{U_l(X)|\forall l \in L, \forall t \in T\},
\]

(2)

where \( N \) is the maximal amount of samples required at each location.

According to [14], the QoI satisfaction ratio \( S(X) \) by selecting \( X \) can be denoted as:

\[
S(X) = 1 - \sqrt{\frac{\sum_{m \in X} |\forall l \in L, \forall t \in T (N - U_l(X))\}}{L \cdot T}.
\]

(3)

The goal of the participant coordination procedure is to maximize \( U(X) \). We choose to borrow (3) because it considers not only the amount, but also the distribution of collected data. When the same amount of data are collected, \( U(X) \) gives high values to uniformly distributed data.

2.2.2. Incentive model

We let \( c_m, \forall m \in M \) denote the incentive requirement of participant \( m \) for his/her data contribution. Therefore, the final incentive \( C(X) \) given to all selected participants can be calculated as:

\[
C(X) = \sum_{m \in X} c_m.
\]

(4)

Here, we further denote the total amount of data uploaded by participant \( m \) as \( a_m = \sum_{l \in L, t \in T} u_m^l |\forall m \in X \). The total amount of collected data can be denoted as:

\[
A(X) = \sum_{m \in X} a_m.
\]

(5)

2.2.3. Privacy metric

We follow the definition of the privacy metrics proposed in [29], which is based on the entropy theory. Entropy is a measure of unpredictability in information theory. Consider that we have \( A \) pieces of uploaded data tagged with locations, privacy is maximized when the attacker sees all participants with equal probability of reporting a piece of data. In this way, the attacker will not be able to identify the owner of the data. Therefore, the degree of anonymity depends on the distribution of the probabilities for the \( A \) pieces of reported data. Let \( P(X) \) denote the privacy provided by selecting \( X \). We have:

\[
P(X) = - \sum_{m \in X} \frac{a_m}{A(X)} \log_2 \frac{a_m}{A(X)}.
\]

(6)

In order to select more participants to increase the privacy level, we need to select participants with both lower incentive requirements and less amount of data collection.
Combining (3), the goal of participant coordination is to find a set of participants, \( \mathcal{X} \), that fulfills:

\[
\begin{align*}
\text{max} & : S(\mathcal{X}) \\
\text{max} & : P(\mathcal{X}) \\
\text{s.t.} & : C(\mathcal{X}) \leq B
\end{align*}
\]  

(7)

2.3. Threat model

In order to select the most efficient participants for sensing task, some personal information will be submitted to the application server usually. Thus the risk of privacy leakage will be produced in both sides, the application server and participants.

For the application server side, we consider the application server not trustworthy for protecting participants’ privacy in our model. Any information learned by the application server might be leaked to malicious administration personnel behind the application server. For example, given the trajectories of participants, a malicious attacker can infer sensitive information, such as alternative lifestyles or political affiliations. However, we trust the application server in terms of its functionality, e.g., participant coordination, data correctness calculation, and incentive calculation.

For the participant side, we allow anyone with an appropriate device that gets the application installed to register as a participant. All the participants must use the specified application to reduce the risk of privacy leakage when they collect or upload sensory data. Here we assume that the participants are selfish and rational; On one hand, they are willing to cooperate with others in order to get rewarded; On the other hand, they tend to enlarge their own profits by distorting the data report of others if it is profitable. There can also be misbehaving participants, who produce false sensing data or send false data randomly with certain probability (on-off attacks), but the majority of the reports are good.

Different from existing works that include reliable TTP servers, we do not assume that the registration server is reliable. It means that the application server can be controlled by the adversary. Also we consider that an adversary could pretend himself as an ordinary participant and spy on the passing information.

3. The proposed privacy-preserving mechanisms

As stated above, two mechanisms are designed for the overall procedure of data collection: (1) For participant coordination, the application server selects participants iteratively; (2) for data aggregation, incentive distribution, and punishment mechanism, the participants iteratively pass their collected data tagged with accurate locations through other participants and finally to the application server, and the incentives are sent to participants at the end of the sensing task.

3.1. Participant coordination

We first introduce the general mechanism of participant coordination, especially the communication procedure between application server and participants, followed by the detailed introduction of the participant selection algorithm implemented on the application server.

3.1.1. The general mechanism of participant coordination

The general mechanism is shown in Fig. 2, and the detailed descriptions are also given below:

Step 1. All the potential participants are informed by the application server with the data requirement of the task. In the first iteration, the application server informs all the participants the original task requirement \( R = r_1 = N \forall w \in \mathcal{L} \). However, in the later rounds, each unselected participant queries new data requirements \( R' = R - D_{m_l} \), from the winner \( m_1 \) in the former round according to the given IP address of \( m_1 \).

Step 2. All the potential participants submit their QoI satisfaction ratio increment to the application server. The application server selects a winner each round according to certain judgment function which will be introduced later. If no participant can be selected, the iteration stops.

Step 3. The application server informs the winner \( m_2 \) that he/she is selected in this round. The winner updates the data requirement of the task, \( R' = R - D_{m_2} \).

Step 4. The application server informs the unselected participants the IP address of the winner \( m_2 \). The iteration repeats from step 1.

On one hand, such mechanism can guarantee that each participant knows in real-time about the not-yet-satisfied data collection requirements. Therefore, it can guarantee that the selected participant in each iteration can provide optimal QoI contribution and avoid redundant data collected by multiple participants. On the other hand, each participant only needs to reveal his/her location to nearby participants, so that the application server will not be able to collect the trajectories of the participants.

3.1.2. The proposed participant selection strategy

Here we further discuss what statistical information the participants should upload to the application server, and what criteria the application server uses to select the winner.

If the trajectories of all participants are known as a prior, we can easily find an optimal set of \( \mathcal{X} \) or a suboptimal set of \( \mathcal{X} \) by optimization algorithms. In our proposed mechanism, participants do not upload their trajectories to the application server, but the trajectories can be easily restored if the location information of the uploaded sensory data are very sparse. Thus, we adopt an approximation algorithm to solve this problem. By observing (7), we find that selecting participant \( m \) who (1) contributes more to the overall QoI satisfaction ratio; (2) has lower incentive requirement \( c_{m_i} \); and (3) has smaller amount of data collection \( a_{m_i} \) can help to (1) achieve the optimization goal; (2) lessen the total incentive requirement; and (3) improve the privacy metric, respectively.
To achieve the purpose, we use an iterative participant selection strategy. The steps of the proposed participant selection strategy is as follows. First, we use a Borda Count [37] based algorithm to calculate the total weighted contribution of a participant with his/her data contribution and incentive requirement. The Borda count is an election method in which voters rank options or candidates in order of preference. The Borda count determines the outcome of a debate or the winner of an election by giving each candidate, for each ballot, a number of points corresponding to the number of candidates ranked lower. Once all votes have been counted the option or candidate with the most points is the winner. The benefit of using Borda count is that it can select broadly acceptable candidates, rather than those preferred by a majority. In other words, it can best balance the requirements of higher QoI satisfaction ratio increment, less amount of data collection and lower incentive requirement. Second, we find the participant associated with the highest value. Third, we select participants one by one until the QoI requirements of all tasks are fully satisfied, or until the task budget runs out. In such an iterative way, the proposed algorithm selects the most “efficient” participants.

The algorithm is presented in pseudocode by Algorithm 1, and is described in details below:

Step 1. **Initializing:** Suppose at the beginning of an iteration step, a set of participants, $X$, have already been selected. For an unselected participant $m$, he/she should report to the application server three statistical values: (1) the QoI satisfaction ratio increment brought by him/her $s_m = S(X + m) - S(X)$; (2) his unit incentive requirement $c_m$; and (3) the total amount of his data collection expectation $a_m$.

$$k_m = \alpha \cdot k_m^i + \beta \cdot k_m^c + \gamma \cdot k_m^a$$

where $\alpha$, $\beta$, and $\gamma$ denote the weights of each ranking. If we have a preference for one of these values, we can increase its weight. Here we set all the three weights to 1.

3.2. Confusion mechanism

When the process of participant selection is complete, due to the winners in each iteration sent their utility to...
Algorithm 1 Proposed dynamic participant selection scheme.

Require:

unselected participants $m_1, m_2, m_3, \ldots, m_n (n > k)$; number of candidate participants $k$; array of QoI satisfaction ratio increment of each participant $s[m_1] \ldots s[m_n]$; array of unit incentive requirement of each participant $c[m_1] \ldots c[m_n]$; array of data amount of each participant $a[m_1] \ldots a[m_n]$; the set of selected participants $\chi'$.

Ensure: Selected participant of this step of iteration $m^*$;

1: initial array $s'$, array $c'$, array $a'$, array rank;
2: $s' = \text{sort}(s)$;
3: $c' = \text{sort}(c)$;
4: $a' = \text{sort}(a)$;
5: selected = NULL;
6: for participant $m \in m_1, m_2, m_3, \ldots, m_n$ do
7:     for each $s'$ as $j => i$ do
8:         if $m == i$ then
9:             rank[$m'] += $i \cdot \alpha$;
10:        end if
11:    end for
12:    for each $b'$ as $j => i$ do
13:        if $m == i$ then
14:             rank[$m'] += $i \cdot \beta$;
15:        end if
16:    end for
17:    for each $a'$ as $j => i$ do
18:        if $m == i$ then
19:             rank[$m'] += $i \cdot \gamma$;
20:        end if
21:    end for
22: end for
23: sort(rank);
24: selected = rank(1);
25: for tmp in 1, 2, ..., $k$ do
26:     compute the incentive $C$ of $(\chi' + \text{rank[ tmp ]})$ in (4)
27: if $C \leq B$ then
28:     selected = tmp;
29:     break;
30: end if
31: end for
32: Return: final selected participant selected.

3.3. Data aggregation, incentive and punishment mechanism

In the participant coordination mechanism, each selected participant has a new IP address as a sensory data sender in the data transfer sequence, and the participants have been informed with the participants before and after him/her, thus a participant knows where the sensory data will be received and sent to. We call the participant before this participant as his/her parent, and the participant after as his/her child. The procedure of data aggregation, incentive, and punishment mechanism is described in details below:

Step 1. First participant: When task time completes, the first participant in the data transfer sequence will send the value of obtained QoI satisfaction ratio to the application server and send the sensory data to his/her child.

Step 2. Iterative participant: In the iterative procedure of data aggregation, each participant first receives sensory data from his/her parent, then he/she calculates the QoI satisfaction ratio before and after he/she adds his/her sensory data into the aggregative data set, and he/she will send the two value of QoI satisfaction ratio to the application server and send the aggregative data set to his/her child.

Step 3. Last participant: After the last participant receives the aggregative data set from his/her parent, he/she calculates the QoI satisfaction ratio before and after he/she adds his/her sensory data into the aggregative data set, and he/she will send the two value of QoI satisfaction ratio and the aggregative data set to the application server.

Step 4. Application server: When the application server receives the final aggregative data set, the application server will contrast each QoI satisfaction ratio sent by a participant and his/her child (the last QoI satisfaction ratio will be calculated by the application server). These two values should be equal, but if there is an abnormality, the application server will connect the relevant participants and make the judgment and punish the party at fault. Then the application server will send the final payments to every participant.

The methods of punishment include but are not limited to reducing incentive rewards, reducing reputation, or adding participants to the blacklist, etc. In the whole process, we consider the corresponding risk from participants for participant selection, data collection, and aggregation: (1) in participant selection, the participants may submit a fabricated QoI satisfaction ratio increment for higher incentive; (2) in data collection, the participants may use fabricated data for “saving time”; (3) in data aggregation, the participants may forget to send data or the data transfer may be failed due to the network problem etc.

For problem 1, in our proposed method, the value of obtained QoI satisfaction ratio by a participant will be send to the application server by himself/herself, and his/her child also calculates it and send it to the application server, thus if a participant send a fabricated QoI satisfaction ratio increment to the application server, he/she must be others, a part of their future location information may be deduced, thus the participants still face the risk of privacy leaks. To solve this problem, a confusion mechanism will be used in the following data collection and aggregation.

After the application server selects the last participant, the application server will get the iterated array of selected participants, then the application server will randomly shuffle the array. The new array will be used as the data transfer sequence.
punished later. And if the child give the fabricated value, the application server can also make the judgment by calculating their obtained sensory data and punish the child.

For problem 2, the application server cannot distinguish the sensory data, thus the application server could not prevent this from happening, but the application server let the participants use a special software application for data collection. The special software application only allow participants to collect and transfer sensory data, and the software application could also record their behaviors, such as their submitted QoI satisfaction ratio increment, etc., thus the participants could not use it to cheat.

For problem 3, if the data transfer is interrupted for a long time (exceeding the threshold), due to some participant has sent the value of obtained QoI satisfaction ratio to the application server, the application server could easily find where the problem happens (the participant who reported lastly or his/her child), then the application server can connect them to solve the problem. If the child does not receive the sensory data, the application server will let the parent resend it; if the child is disconnected, the application server will let the parent send the sensory data to the next node (the child’s child) and punish the child later.

4. Privacy analysis

We summarize the following mandatory attributes for a privacy-preserving participant coordination mechanism, as well as its related data aggregation and incentive distribution operations:

(A) The application server cannot know the future trajectory of any participant, nor associate a piece of collected data with a particular participant.

(B) Multiple sensing reports from the same participant are not linkable.

(C) A participant is rewarded according to his/her required incentive, and cannot benefit from uploading less or false sensory data.

Here we analyze and prove that the proposed mechanism can achieve attributes (A)–(C), and the mechanism itself is secure.

Proposition 1. The application server cannot know the trajectory of a participant during participant coordination. (A)

According to (3), the participants are only uploading the QoI satisfaction ratio increment, but not the trajectories itself. The same QoI satisfaction ratio increment can be achieved by multiple trajectories, so the application server cannot know the future trajectory of a participant.

Proposition 2. The application server cannot see the participant ID from a sensing report, or associate a piece of collected data with a particular participant. (A)

In data aggregation procedure, application server receives a number of sensing reports, collected by multiple participants. Therefore, the possibility of distinguishing the identity of the collector of a certain piece of data at a certain location is low.

Proposition 3. The application server cannot correlate multiple sensing reports of the same participant. (B)

All the sensing reports are aggregated together as one, so that when the application server gets the sensing reports, it could not discriminate them.

Proposition 4. Misbehaving participants will get less rewarded by sending less sensory data, or be punished by sending false sensory data. (C)

The software application will record the participants’ data utility when they submit the QoI satisfaction ratio increment and make a judgment if the collected data can satisfy their submitted the recorded data utility, and these situations will be sent to the application server. Thus, the application server can finally decide if a participant should be punished.

On the other hand, in the stage of data aggregation, as the sent sensory data of participants will be verified by their children, when a participant sends less sensory data to his/her child, if the child does not verify it, the child will sent less sensory data to his/her child and may be punished by the application server. Therefore, in order to guarantee respective rights, each participant should observe the rules.

5. Evaluation, results, and discussions

5.1. Setup

We evaluate the proposed scheme by simulations using [38], the dataset of mobility traces of taxi in Rome, Italy. It contains GPS coordinates of approximately 320 taxis collected over 30 days. Each trajectory is marked by a sequence of time-stamped GPS points that contain taxi driver id, time stamp (date and time), and taxi drivers’ position (latitude and longitude). We adopt the following procedures to set up our simulation platform: We adopt the following procedures to set up our simulation:

- As all traces were spread in different parts of Rome, the number of GPS points are too big (more than twenty million) and the traces are distributed very unevenly, as shown in Fig. 3(a), so a specific rectangular region where the traces mostly appear is needed. We stored all traces and found a region about 800 m × 500 m as shown in Fig. 3(b). This region is closed to Parco Adriano, a park in Rome, on the northern bank of the Tiber, just to the east of the Vatican. We use this region as the simulation area for the considered data collection application. Fig. 3(c) and (d) show the GPS points and traces.

- In order to simplify the computation in our experiments, the entire region was divided into 16 × 10 sub-regions of 50 m × 50 m, i.e., L = 160. For each location, the maximal required amount of data was set to be 100 (N = 100). Since a participant’s incentive requirement could be realized in different formats in practice, such as real money or bonus points, we use dimensionless units to represent both the participants’ incentive requests and the tasks’ budget constraints. The default budget of sensing task was set to be 1000, or:
C = 1000, the incentive requests for participants were set to random numbers from 1 to 10.

- All the 1040 traces in the considered region were taken as potential (candidate) participants, i.e., M = 1040. Since these traces were recorded at different times, in our simulation we simply neglected their timestamps and overlaid them into the same time period. Fig. 3 shows all of the 1040 traces. The length of these traces are different, most of them are from 50 to 200.

- All experiments had been repeated at least a hundred times (more for random selection method), and we took the average values as the final results.

We compare our proposed participant coordination scheme, referred as “BPS” in all figures, with the random selection method that randomly selects participants till the incentive budget is exhausted. Besides, we also compare with the QoI optimal solutions, which keep selecting participants one at a step in a greedy way, according to their QoI contributions and their incentive requirements, respectively, until the incentive budget runs out. All these four schemes and environmental settings are written by script files in Matlab.

5.2. Results and discussions

First we investigate the impact of task incentive budget C on the attained QoI satisfaction ratio. From Fig. 4(a) we observe that when the incentive budget is limited, the QoI optimal method can provide a little better QoI satisfaction retro than our proposed method, but the gap between them is reducing with the growth of the incentive budget. When the budget is limited, selecting participants with lower incentive requirements can significantly increase the amount of data collection. On the other hand, when the budget is adequate, it becomes more important to consider whether the data collection capability of a participant fits the data requirement of a task. Nevertheless, our proposed method can always achieve better QoI satisfaction ratio and privacy level than random selection method.

As shown in Fig. 4(b), we can observe that the percentage of redundant data is lower than others if using our proposed method especially when the incentive budget is limited. With the growth of incentive budget, the value of our proposed method improving faster, gradually close to the value of the QoI optimal method, but still lower than the random selection method.

Moreover, we can observe that with the growth of incentive budget, our proposed method will obtain higher...
Fig. 4. Simulation results of (a) attained QoI satisfaction ratio, (b) percentage of redundant data, (c) attained privacy level when varying the incentive budget, and (d) number of selected participants when varying the task incentive budget.

privacy level. However, when more budgets are provided, our proposed method starts to perform much better, as shown in Fig. 4(c). The growth rate of QoI optimal method is changing more gently shows that the growth of attained QoI satisfaction ratio is increasingly difficult even if the budget is adequate, thus if more attained QoI satisfaction ratio is required, the participants will face a higher risk of privacy leaks. We also observe that our proposed method is much better than random selection method. Although the obtained redundant data are less than other compassion methods, our proposed method could select more participants due to consider the participant privacy, as shown in Fig. 4(d).

Then we investigate the impact of the number of candidate participants $M$. As shown in Fig. 5(a), we can observe that when the number of candidate participants is limited, almost all the candidate participants are selected, so that the attained QoI satisfaction ratio and privacy level are very close for different methods. With the growth of the number of candidate participants, due to the different amount of uploaded sensory data for participants, the attained QoI satisfaction ratio of random participant selection even declines. Our proposed method is much better than random participant selection, and the attained QoI satisfaction ratio is very close to the QoI optimal method.

When the number of candidate participants is limited, the attained QoI satisfaction ratio is also very limited and nearly all candidate participants will be selected, so the percentage of redundant data is not very high and very closed to each other, as shown in Fig. 5(b). And with the growth of the number of candidate participants, if there are more participants could be selected, the application server could select more efficient participants that help to both improve QoI satisfaction ratio and reduce the percentage of redundant data. However, the value of our proposed method is always higher than other compassion methods.

And similar with Fig. 4(c), the privacy level of our proposed method is getting better and better than the QoI optimal method when more participants could be selected. Meanwhile, the number of selected participants is also getting larger and larger of our proposed method in order to improving the participant privacy, as shown in Fig. 4(d).

Finally, we investigate the impact of the required amount of data in each subregion $N$. As shown in Fig. 6(a), we can observe that with the growth of required amount of data in each subregion, the attained QoI satisfaction ratio of each method keeps reducing, but the value of our proposed method is closed to the QoI optimal method and much higher than the random selection method. However, due to more sensory data are required, the percentage of redundant data also reduces with the growth of required amount of data, and the speed of our proposed method is faster, as shown in Fig. 6(b).

Meanwhile, due to the task incentive and the number of candidate participants are not changed, the privacy level and the number of selected participants could not change too much, as shown in Fig. 6(c) and (d). However, the

Please cite this article as: B. Zhang et al., Privacy-preserving QoI-aware participant coordination for mobile crowdsourcing, Computer Networks (2016), http://dx.doi.org/10.1016/j.comnet.2015.12.022
Fig. 5. Simulation results of (a) attained QoI satisfaction ratio, (b) percentage of redundant data, (c) attained privacy level when varying the incentive budget, and (d) number of selected participants when varying the number of candidate participants.

Fig. 6. Simulation results of (a) attained QoI satisfaction ratio, (b) percentage of redundant data, (c) attained privacy level when varying the incentive budget, and (d) number of selected participants when varying the required amount of data in each subregion.

Please cite this article as: B. Zhang et al., Privacy-preserving QoI-aware participant coordination for mobile crowdsourcing, Computer Networks (2016), http://dx.doi.org/10.1016/j.comnet.2015.12.022
values of our proposed method are still much better than other comparison methods.

6. Conclusion and future work

In this paper, a privacy-preserving participant coordination mechanism is proposed to both achieve optimal QoS for sensing tasks, and protect the location privacy of participants. Specifically, the cooperation among participants are used to replace the traditional centralized participant coordination phase. An optimization problem is formulated to select participants iteratively to maximize the QoS satisfaction ratio and privacy level while fulfilling the constraints of incentive. Based on this, an approximate solution is proposed based on Borda Ranking method. A cooperative data aggregation method is further proposed to protect participant privacy through the whole data collection procedure. And a punishment mechanism is proposed to ensure continuous operation of the whole system. Extensive simulation results, based on a real trace dataset of taxi in Rome, showed the effectiveness and robustness of our approach.

In the future, we plan to further consider the reputation of participants for our punishment mechanism to ensure the data quality. Extensively, we will consider not only the privacy protection issue but also the security issues in such systems. For example, we would like to investigate how to protect the privacy of participants when the registration server pretends itself as a smart device to spy on the trajectory information of other participants. Meanwhile, we plan to explore how to further strengthen the privacy protection strategy to prevent privacy leaks in the process of information exchange between participants and server cloud.

Acknowledgments

This paper was financially sponsored in part by Prospective Research Project on Future Networks under grant no. BY2013095-2-18, the National Natural Science Foundation of China (grant no. 61301079, 12671041, and 61371017), and in part by the National High Technology Research and Development Program of China (863 Program, no. 2013AA013301).

References


Please cite this article as: B. Zhang et al., Privacy-preserving Qoi-aware participant coordination for mobile crowdsourcing, Computer Networks (2016), http://dx.doi.org/10.1016/j.comnet.2015.12.022
Bo Zhang received the M.S. degree from Beijing University of Posts and Telecommunications, China, in 2010, and currently he is a Ph.D. student with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, China. Prior to studying in BUPT, he worked in the R&D Department, H3C Technologies Co., Ltd. His research interests include participatory sensing and the Internet of Things (IoT).

Chi Harold Liu (M’10–SM’15) receives his Ph.D. degree in electronic engineering from Imperial College, London, U.K., in 2010, and his B.Eng. degree in electronic and information engineering from Tsinghua University, Beijing, China, in 2006, respectively. He is currently a full professor and serving as the assistant dean at the School of Software, Beijing Institute of Technology, China. From 2014, he also serves as the Director of Data Science Institute, the Director of IBM Mainframe Excellence Center (Beijing), the Director of IBM Big Data and Analysis Technology Center, and the Director of National Laboratory of Data Intelligence for China Light Industry. Before moving to academia, he worked for IBM T. J. Watson Research Center and IBM Research—China as a staff researcher and project manager from 2010 to 2013, worked as a postdoctoral researcher at Deutsche Telekom Laboratories, Germany in 2010, his current research interests include the Internet of Things (IoT), big data analytics, mobile computing, and wireless ad hoc sensor, and mesh networks. Prof. Liu was elected into the “High-Level Overseas Talents Return Home Program”, by Ministry of Human Resource and Social Security, China, in 2015, received the Distinguished Young Scholar Award from Beijing Institute of Technology in 2013, received IBM First Plateau Invention Achievement Award in 2012, received IBM First Patent Application Award in 2011, and was interviewed by EEWeb.com as the Featured Engineer in 2011. He has published more than 80 prestigious conference and journal papers and owned more than 10 EU/US/China patents. He serves as the editor for KSI Transactions on Internet and Information Systems from 2013, and the book editor for six books published by Taylor & Francis Group, USA and Wiley. He also has served as the general chair of IEEE SECON’13 workshop on IoT Networking and Control, IEEE WCNC’12 workshop on IoT Enabling Technologies, and ACM Ubicomp’11 Workshop on Networking and 3D Object Memories for IoT. He served as the consultant to the Asian Development Bank, Bain & Company, and KPMG, USA, and the peer reviewer for Qatar National Research Foundation, Qatar, and National Science Foundation, China. He also serves as the Lead Guest Editor for IEEE Sensors Journal, Special Issue on Software Defined Wireless Sensor Networks, and Guest Editor for IEEE Transactions on Emerging Topics in Computing. Special Issue on Sensor Data Computing as a Service in Internet of Things. He is a senior member of Chinese Institute of Electronics, a senior member of IEEE, and a member of ACM. For more information, please visit: http://haroldliu.weebly.com

Jianyu Lu is an undergraduate student at School of Optic and Electronic Information Engineering, Huazhong University of Science and Technology. In 2015 Summer, he worked as a data analyst at BIM, Baidu Online Network Technology Co. Ltd. His research interests include Internet of Things, mobile cloud, emotion-aware system, and data analytics, etc.

Zheng Song is a Ph.D. student with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, China. Prior to studying in BUPT, he worked in the R&D Department, Sina Corp and owned 10 U.S./China patents. His research interests include participatory sensing, indoor localization and the Internet of Things (IoT).

Ziyu Ren is an undergraduate student in the School of Information Science and Technology, Tsinghua University, since 2014. He was a gold medal winner of the 30th Chinese National Youth Information Science Olympiad. His research interests include mobile crowdsourcing, social networks, and the Internet-of-Things (IoT), etc.

Jian Ma received the B.Sc. and M.Sc. degrees from BUPT in 1982 and 1987, respectively, and Ph.D. degree in 1994 from the Department of Electronics Engineering, Helsinki University of Technology, is a principal member of Research Staff of Nokia Research Center, Beijing. He is also a guest professor in Computer Science, Beijing University of Posts and Telecommunications (BUPT) from 2002, IET-CAST from 2003, and CCCE-GUCAS from 2006, and now supervising about one and half dozens of Ph.D. and M.Sc. students at BUPT, CCCE-GUCAS, IET-CAST, and Tsinghua University, China. He authored more than 160 conference and journal papers, and 4 books and book chapters. He has been involved in few professional activities, currently serving a Council Member of Beijing Communication Institute (a senior consultant club in communication industry and academia), and a vice chairman of Sensor Network Technical Committee, China Computer Federation. He has also been serving numerous co-chairs and members of organization committees or program committees in several leading conferences like ICC, AINA, PIMRC, ICDCS, WWW, ChinaCom, etc.

Wendong Wang received both of his bachelor and master degrees from Beijing University of Posts and Telecommunications, China, in 1985 and 1991, respectively, where he is currently a full professor. He is currently serving as the head of the National Natural Science Foundation program and the National High Technology Research and Development Program of China. He has published more than 200 papers in various journals and conference proceedings, and owned 14 U.S./China patents. His current research interests include the next generation network architecture, Internet of Things, participatory sensing, wireless ad-hoc, sensor and mesh networks, and mobile internet. He is a member of ACM and IEEE.