Privacy-Preserving Data Aggregation Framework for Mobile Service Based Multiuser Collaboration

Hai Liu[1], Zhenqiang Wu[1], Changgen Peng[2], Feng Tian[1], and Laifeng Lu[3]
[1]School of Computer Science, Shaanxi Normal University, China
[2]Guizhou Provincial Key Laboratory of Public Big Data, Guizhou University, China
[3]School of Mathematics and Information Science, Shaanxi Normal University, China

Abstract: *Considering the untrusted server, differential privacy and local differential privacy has been used for privacy-preserving in data aggregation. Through our analysis, differential privacy and local differential privacy cannot achieve Nash equilibrium between privacy and utility for mobile service based multiuser collaboration, which is multiuser negotiating a desired privacy budget in a collaborative manner for privacy-preserving. To this end, we proposed a Privacy-Preserving Data Aggregation Framework (PPDAF) that reached Nash equilibrium between privacy and utility. Firstly, we presented an adaptive Gaussian mechanism satisfying Nash equilibrium between privacy and utility by multiplying expected utility factor with conditional filtering noise under expected privacy budget. Secondly, we constructed PPDAF using adaptive Gaussian mechanism based on negotiating privacy budget with heuristic obfuscation. Finally, our theoretical analysis and experimental evaluation showed that the PPDAF could achieve Nash equilibrium between privacy and utility. Furthermore, this framework can be extended to engineering instances in a data aggregation setting.*

Keywords: *Differential privacy, Nash equilibrium, conditional filtering noise, adaptive Gaussian mechanism, PPDAF.*

Received November 22, 2017; accepted October 4, 2018
https://doi.org/10.34028/iajit/17/4/3

1. Introduction

Nowadays, the development and applications of mobile devices generate massive data. Third server needs to aggregate these data to provide service to mobile users. But this could lead to privacy concerns for users, because untrusted third party server analyzes data or sells data for benefit. Privacy is one of the most important properties of an information system must satisfy to share information among different untrusted entities, and the protection of sensible information plays an important role [8]. On the assumption that the server is reliable, so Dwork *et al*. [5] proposed differential privacy. In real world, using differential privacy may lead to sensitive information leakage of centralized data because the third server is not reliable by analyzing or abusing users' data. Thus, differential privacy has used to privacy-preserving by user adding noise to own data in data aggregation setting.

Exponential mechanism can protect the data privacy of participants to untrusted third party and encourage players to honestly report information [18]. Moreover, local differential privacy [12] has been provided for privacy-preserving in a local setting. Local differential privacy can be achieved by using randomized response [26] technology with providing plausible deniability for users responding to sensitive surveys. However, differential privacy and local differential privacy cannot ensure Nash equilibrium between privacy and utility to mobile service based multiuser collaboration. Multiuser collaboration is multiuser negotiating a desired privacy budget in a collaborative manner to send noise data to a server in data aggregation setting.

Therefore, we present an adaptive Gaussian mechanism maintaining Nash equilibrium between privacy and utility, and apply it to Privacy-Preserving Data Aggregation Framework (PPDAF). We show that our framework achieves Nash equilibrium between privacy and utility. Moreover, our framework is easy to extend to engineering implementation for a data aggregation setting. The contribution of this paper can be summarized as follows.

- We analyzed differential privacy and local differential privacy no satisfying Nash equilibrium between privacy and utility for mobile service based multiuser collaboration in a data aggregation setting.
- We gave adaptive Gaussian mechanism ensuring Nash equilibrium between privacy and utility by multiplying expected utility factor with conditional filtering noise under expected privacy budget.
- We constructed PPDAF using adaptive Gaussian mechanism based on negotiating privacy budget with heuristic obfuscation, which kept Nash equilibrium between privacy and utility.

The rest of this paper is organized as follows. Section 2 introduces related work. Section 3 introduces the preliminaries. Section 4 gives equilibrium analysis of differential privacy and local differential privacy for mobile service based multiuser collaboration. Section 5 presents adaptive Gaussian mechanism. Section 6 details PPDAF. Section 7 conducts experimental and comparison analysis. Section 8 concludes this paper.

2. Related Work

To secure against malicious participants, Dwork *et al.* [6] provided efficient distributed protocols for generating shares of random noise by using verifiable secret sharing. To prevent the privacy leakage of participants to untrusted third party and encourage players to honestly report information, McSherry and Talwar [18] proposed exponential mechanism. Secure function evaluation is a natural paradigm for distributed differential privacy analysis [1].

In distributed data aggregation, differential privacy had been applied to continual monitoring of heavy hitters from distributed streams against an untrusted aggregator [2] and data collection for untrusted mobile crowdsensing [20]. For multi-party computation, McGregor *et al.* [15] studied differential privacy computation for which the two parties would like to perform analysis of their joint data while preserving privacy of both datasets. The recommendation service with collective user behaviour using differential privacy was studied [17]. Feild *et al.* [9] described an approach to distributed search log collection, storage, and mining, with the dual goals of preserving privacy and making the mined information broadly available. Since privacy leakage of data releasing for multi-party environment, Mohammed *et al.* [19] presented a two-party protocol for the exponential mechanism. Su *et al.* [25] studied the problem of publishing high-dimensional data in a distributed multi-party environment using differential privacy. For big data analytics in the distributed setting, Li *et al.* [13] developed a privacy-preserving distributed online learning framework by introducing the notion of differential privacy on the data collected from distributed data sources. Shokri and Shmatikov [23] proposed a practical system based on differential privacy that enabled multiple parties to jointly learn an accurate neural-network model for a given objective without sharing their input datasets. Differential privacy also had applied to collaborative search log [10] and crowdsourced spectrum sensing [11] to prevent an untrusted third party from learning user's privacy information.

Kasiviswanathan *et al.* [12] first formalized the local privacy model. Duchi *et al.* [4] proposed local differential privacy in which data remains privacy even from the statistician or learner. Since local differential privacy algorithms were highly interactive, Smith *et al.* [24] provided new algorithms which were either noninteractive or use relatively few rounds of interaction. Local differential privacy had got more applications, such as local privacy hypothesis testing [21]. Also, Cormode *et al.* [3] introduced the key technical underpinnings of deployed systems using local differential privacy of Google, Apple and Microsoft.

In summary, differential privacy and local differential privacy have been used for data aggregation environment. But differential privacy and local differential privacy cannot ensure Nash equilibrium between privacy and utility. Therefore, we proposed an adaptive Gaussian mechanism and applied it to a data aggregation environment.

3. Preliminaries

We introduce the preliminaries to Nash equilibrium and differential privacy.

3.1. Nash Equilibrium

Nash equilibrium [22] is a stable solution concept, which all players can achieve expected utility using best strategy response.

- *Definition* 1 *(Nash Equilibrium)*: In a game, a strategy profile $s=(s_1,\ldots,s_n)$ is a Nash equilibrium, when the utility function $u_i(s_i, s_{-i}) \geq u_i(s_i^*, s_{-i})$ of any strategy $s_i^* \in S_i$ of every player i.

Here, $s_{-i}=(s_1,\ldots,s_{i-1},s_{i+1},\ldots s_n)$ denotes a strategy profile s without player i's strategy.

3.2. Differential Privacy

We present differential privacy and it corresponds to mechanisms [7]. Two datasets D_1 and D_2 are adjacent datasets, when Hamming distance $d(D_1,D_2)$ is 1.

- *Definition* 2 *((ε,δ) -Differential Privacy)*: Given $\varepsilon \geq 0$, a randomized algorithm M is (ε,δ)-differential privacy, if for adjacent datasets D_1 and D_2 and for any outputs $S \subseteq Range(M)$ of M, then

$$\Pr[M(D_1) \in S] \leq e^{\varepsilon} \Pr[M(D_2) \in S] + \delta \qquad (1)$$

Where the algorithm M is ε-differential privacy with probability at least $1-\delta$. If $\delta = 0$, M is $(\varepsilon,0)$-differential privacy algorithm.

Differential privacy has the property of parallel composition [16].

- *Theorem* 1 *(Parallel Composition)*. Each random mechanism M_i provides (ε_i,δ)-differential privacy. The D_i be arbitrary disjoint subsets of the input dataset D. The parallel composition of M_i is $(\max\{\varepsilon_i\},\delta)$-differential privacy.

For any query function $f: D \rightarrow R^k$ of a dataset D, the sensitivity of f is

$$\Delta f = \max_{d(D_1,D_2)=1} \| f(D_1) - f(D_2) \|_1 \qquad (2)$$

for all adjacent datasets D_1 and D_2. R is the set of all real numbers.

- *Definition* 3 *(Gaussian Mechanism)*: Given any query function $f: D \to R^k$, the Gaussian mechanism is defined as $GM(D)=f(D)+Y$, where $Y=\{Y_1,\dots,Y_k\}$ is independent identical distribution random noise drawn from Gaussian distribution $N(0,\sigma^2)$ and $\sigma \geq (\Delta f\sqrt{2\ln(1.25/\delta)})/\varepsilon$.

Considering two input values of individuals, the definition of local differential privacy [12] is as follows.

- *Definition* 4 *(Local Differential Privacy)*: Given $\varepsilon \geq 0$, a randomized algorithm M is ε-local differential privacy, if for any input b_1 and b_2 and for any output $v \in \{v_1, v_2\}$, then

$$\Pr[v \mid b_1] \leq e^{\varepsilon} \Pr[v \mid b_2] \tag{3}$$

4. Equilibrium Analysis of Differential Privacy and Local Differential Privacy

- *Theorem* 2. Differential privacy cannot achieve Nash equilibrium between privacy and utility for mobile service based multiuser collaboration.
- *Proof*: A single user can achieve expected privacy budget ε by adding nose to own data, but the utility of aggregated data is destroyed. If the server wants the utility of aggregated data to be U, then the user cannot obtain expected privacy budget. Thus, user and server cannot reach Nash equilibrium between expected privacy budget and expected aggregated data utility for mobile service based single user.

Next, there are n users (every user $i \in \{1,\dots,n\}$) and every user gets expected privacy budget ε_i by adding noise to own data. Firstly, if n users negotiate a privacy budget ε_0 and $\varepsilon_0 \leq \varepsilon_i$ for every user i, then all users obtain desired privacy budget in a distributed setting. However, this leads utility disaster of aggregating data for the server. Secondly, if the expected aggregated data utility of server is U, then some users cannot achieve expected privacy-preserving for their own data. Thus, users and the server cannot achieve Nash equilibrium between expected privacy budget and expected aggregated data utility for mobile service based multiuser collaboration.

- *Theorem* 3. Local differential privacy cannot achieve Nash equilibrium between privacy and utility for mobile service based multiuser collaboration.
- *Proof*: When use local differential privacy, we use the expectation as utility. When the probabilities of output v_1 and v_2 are $e^{\varepsilon}/(1+e^{\varepsilon})$ and $1/(1+e^{\varepsilon})$ on input value b_1, respectively. The probabilities of output v_1 and v_2 are $1/(1+e^{\varepsilon})$ and $e^{\varepsilon}/(1+e^{\varepsilon})$ on input b_2, respectively. And the probabilities of input values b_1 and b_2 are p and 1-p. Given p and 1-p, the expectation of output values v_1 and v_2 is $((pe^{\varepsilon}+1-p)v_1+(p+(1-p)e^{\varepsilon})v_2)/(1+e^{\varepsilon})$. This proof is similar to the proof of Theorem 3. Thus, we can prove that local differential privacy cannot achieve Nash equilibrium between privacy and utility for mobile service based multiuser collaboration.

To sum up, differential privacy and local differential privacy cannot achieve Nash equilibrium between privacy and utility for mobile service based multiuser collaboration. Thus, we need an adaptive differential privacy framework ensuring Nash equilibrium between privacy and utility in a local setting of mobile service based multiuser collaboration.

5. Adaptive Gaussian Mechanism

In the section, we present the adaptive Gaussian mechanism [14] and analyze its Nash equilibrium.

For any numeric query function f:D→R^k, differential privacy mechanisms generate noise directly added to query outcomes. But the noise may be very large that leads data is not available, or too small to preserve individual's sensitive information. So we construct adaptive differential privacy maintaining Nash equilibrium between privacy and utility. To present adaptive Gaussian mechanism, we give definition of conditional filtering noise.

- *Definition* 5 *(Conditional Filtering Noise)*: The conditional filtering noise $Y = X - k\sigma$ satisfies $|Y| \in (0.5, 1.5)$, and noise X is generated by Gaussian mechanism.

Note that σ is the scale parameter of Gaussian distribution. Assuming $\Phi(u)$ is distribution function of $u \sim N(0,1)$. The probability of a standard normal variable $U = u$ is obtained by querying the standard normal distribution function table.

- *Theorem* 4. Probability of conditional filtering noise $|Y| \in (0.5, 1.5)$ is

$$\Phi\left(k-\frac{0.5}{\sigma}\right)-\Phi\left(k-\frac{1.5}{\sigma}\right)+\Phi\left(k+\frac{1.5}{\sigma}\right)-\Phi\left(k+\frac{0.5}{\sigma}\right) \tag{4}$$

- *Proof*: Since conditional filtering noise Y=X-$k\sigma$ satisfies $|Y| \in (0.5, 1.5)$, expectation of Y is - $k\sigma$. The standard normal variable is U=(Y+$k\sigma$)/σ. We can get Pr[U]=Pr[Y]. Because U is subjected to N(0,1), the probability of U is (4).

Next, we define the expected data utility according to the absolute value of relative error $(x-x')/x$, where x' is the approximate value of x.

- *Definition* 6 *(Expected Data Utility)*: *f*(*D*)+*u* is the approximate value of *f*(*D*), where *u*>0 is called as utility factor to achieve expected data utility. The expected data utility to be $U=(|f(D)|-u)/|f(D)|$ if the absolute value $E\in[0,1]$ of relative error is $E=u/|f(D)|$.

According to Definition 6, we can achieve expected data utility, but this is not satisfying differential privacy if we regard the utility factor *u* as noise. So we get noise *Z*=*uY* by multiplying the conditional filtering noise *Y* with the utility factor *u* of the expected data utility. Then we add noise *Z* to query outcome. Probability distribution of noise *Z* corresponding to the probability distribution of the noise *X* satisfies differential privacy. Like this can achieve approximate expected data utility, while satisfying differential privacy.

To define approximate expected data utility, based on a suitable real number $k\in[0,2]$, expected data utility, and expected estimation error, we give the definition of expected privacy budget. That is to say, how much expected privacy budget is needed to achieve the expected estimation error under expected data utility. The Expected Estimation Error (EEE) as privacy metric is

$$EEE=\sum p(Y)\,\|f(D)'-f(D)\|_1=\sum p(Y)\,|Y| \qquad (5)$$

Here, $f(D)'=f(D)+Y$, and Y is noise generated by Gaussian mechanism.

- *Definition* 7 *(Expected Privacy Budget)*: In differential privacy, for a dataset D, the expected privacy budget function $\varepsilon: k\times U\times EEE\rightarrow R$ maps $k\in[0,2]$, expected data utility, and expected estimation error to the expected privacy budget.

Here, $k\in[0,2]$ is required to ensure privacy-preserving monotonicity of adaptive differential privacy. Privacy-preserving monotonicity refers to a privacy metric decreasing with privacy budget increasing.

Based on conditional filtering noise, the utility factor of expected data utility, and the expected privacy budget, definition of approximate expected data utility is as follows. Note that all approximate values are obtained by rounding in this paper.

- *Definition* 8 *(Approximate Expected Data Utility)*: If conditional filtering noise *Y*=*X*-*kσ* of Gaussian mechanism under expected privacy budget satisfied $|Y|\in(0.5,1.5)$ and expected data utility is *U*=1-*E*, then the approximate expected data utility *AU* is

$$AU=(|f(D)|-|Z|)/|f(D)| \qquad (6)$$

Thus, we have the following definition of adaptive Gaussian mechanism based on Definitions 6 and 7.

- *Definition* 9 *(Adaptive Gaussian Mechanism)*: For utility factor *u* of the expected data utility and conditional filtering noise *Y* under expected privacy budget. For any numeric query function *f*:*D*→*R*^{k}, the adaptive Gaussian mechanism is defined as *AGM*(*D*)=*f*(*D*)+*Z*, where $Z=uY$ $(|Y|\in(0.5,1.5))$, *Y*=*X*-*kσ*, and *X* generated by Gaussian mechanism with scale parameter σ under the expected privacy budget.

Next, we analyze adaptive Gaussian mechanism. Similarly the properties of differential privacy [7], adaptive differential privacy satisfies the properties including group privacy, post processing, serial composition. Adaptive Gaussian mechanism also satisfies parallel composition [16]. Moreover, adaptive Gaussian mechanism has the following properties.

- *Theorem* 5. Adaptive Gaussian mechanism is (ε,δ)-differential privacy.
- *Proof*: Since the Gaussian mechanism *M* is (ε,δ)-differential privacy, then

$$|\ln\frac{\exp(-x^2/(2\sigma^2))}{\exp((x+\Delta f)^2/(2\sigma^2))}|\le\varepsilon \qquad (7)$$

In adaptive Gaussian mechanism, conditional filtering noise *Y*=*X*-*kσ* ($|Y|\in(0.5,1.5)$) and *X* generated by Gaussian mechanism under expected privacy budget. Thus, the probability of $Y_i\in Y$ is corresponding to the probability of $X_i\in X$. So the probability of $Z_i\in Z$ is also corresponding to the probability distribution of $X_i\in X$. Since the probability density function of random variable X is $\exp(-x^2/(2\sigma^2))/\sqrt{2\pi}\sigma$, the probability density function of random variable $Z=uY=uX-uk\sigma$ is $\exp(-(z/u+k\sigma)^2/(2\sigma^2))/u\sqrt{2\pi}\sigma$. Thus, we have

$$\begin{aligned}&|\ln\frac{\exp(-(z/u+k\sigma)^2/(2\sigma^2))}{\exp(-(z/u+\Delta f+k\sigma)^2/(2\sigma^2))}|\\ =&|\ln\frac{\exp(-(y+k\sigma)^2/(2\sigma^2))}{\exp(-(y+\Delta f+k\sigma)^2/(2\sigma^2))}|\\ =&|\ln\frac{\exp(-x^2/(2\sigma^2))}{\exp(-(x+\Delta f)^2/(2\sigma^2))}|\\ \le&\varepsilon\end{aligned} \qquad (8)$$

Thus, adaptive Gaussian mechanism is (ε,δ)-differential privacy.

- *Theorem* 6. When expected data utility of adaptive Gaussian mechanism is *U*=1-*E* to absolute value *E* of relative error, approximate expected data utility *AU* of adaptive Gaussian mechanism is approximately equal to expected data utility *U*.

- *Proof*: In adaptive Gaussian mechanism, we have $AU = (|f(D)| - u|Y|)/|f(D)|$ based on the definition of approximate expected data utility $AU = (|f(D)| - |Z|)/|f(D)|$. Since conditional filtering noise $|Y| \in (0.5, 1.5)$ and $E = u/|f(D)|$, the rounding value of $|Y|$ is 1 and $AU \approx U$.

We give Theorem 7 of adaptive Gaussian mechanism maintaining Nash equilibrium between approximate expected data utility and expected privacy budget.

- *Theorem* 7. Adaptive differential privacy maintains Nash equilibrium between approximate expected data utility AU and expected privacy budget ε.
- *Proof*: Since conditional filtering noise satisfying $|Y| \in (0.5, 1.5)$ under any expected privacy budget. Under an expected privacy budget ε, adaptive Gaussian mechanism can achieve expected privacy-preserving level. Since $u|Y| \approx u$ and $AU \approx U$, the approximate expected data utility is approximately equal to expected data utility U of adaptive Gaussian mechanism. Therefore, adaptive Gaussian mechanism maintains Nash equilibrium between the approximate expected data utility AU and the expected privacy budget ε, and the strategy profile (ε, AU) is a Nash equilibrium solution.

According to the definition of adaptive Gaussian mechanism, our scheme not only gains $AU \approx U$, but also achieves expected privacy budget ε, which ensures expected privacy level of sensitive information. Thus, adaptive Gaussian mechanism can maintain Nash equilibrium between approximate expected data utility AU and expected privacy budget ε.

6. Applications Framework Using Adaptive Gaussian Mechanism

We present the PPDAF using adaptive Gaussian mechanism for mobile service based multiuser collaboration and analyze its Nash equilibrium between privacy and utility.

6.1. Privacy-Preserving Data Aggregation Framework

For mobile service based multiuser collaboration, we construct a PPDAF in Figure 1. Firstly, server sends expected data utility U of aggregating data to all users. Secondly, users send noise data to the server using adaptive Gaussian mechanism. Before sending noise data, all users need to negotiate a minimum expected privacy budget ε_0 in a cooperative manner to guarantee every user's expected privacy budget. We call ε_0 as negotiating privacy budget. This framework ensures expected data utility and achieves expected privacy-preserving level of users.

- *Definition 10 (PPDAF)*: PPDAF is a tuple $PPDAF = (P, D, \varepsilon, U, AGM)$, where

1. P is a set of mobile service based users P_i and server S.
2. D is a set of all users' dataset D_i and aggregated dataset D_s of server.
3. ε is a set of users' privacy budget ε_i.
4. U is expected data utility.
5. $AGM(D_i, \varepsilon_i)$ is adaptive Gaussian mechanism for dataset D_i and privacy budget ε_i of user i.

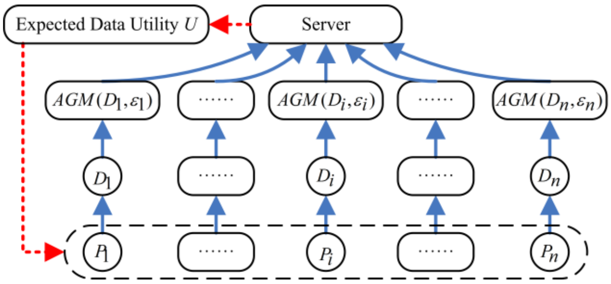

Figure 1. PPDAF using adaptive Gaussian mechanism for mobile service based multiuser collaboration.

Next, we present the principle of negotiating privacy budget ε_0. Since differential privacy guarantees privacy of individuals by adding random noise, the publication of privacy budget does not leak privacy information of individuals. Thus, we construct a heuristic obfuscation of enhancing privacy-preserving for negotiating privacy budget. The process of negotiating privacy budget based on heuristic obfuscation is shown in the Figure 2.

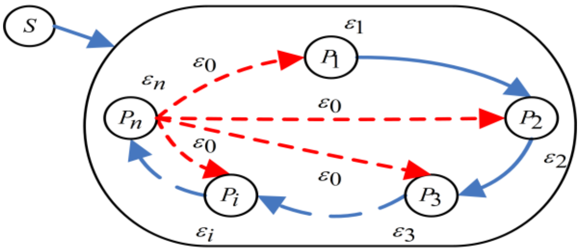

Figure 2. Process of negotiating privacy budget based on heuristic obfuscation for mobile service based multiuser collaboration.

In PPDAF, combining server with users, the process of negotiating privacy budget with heuristic obfuscation is as follows.

- *Step* 1: Server S sets a group $\{P_1, \ldots, P_n\}$ of a mobile service based users of participating data aggregation.
- *Step* 2*:* Server requires all users negotiating privacy budget ε_0 according to the order of $P_1, \ldots, P_n$ based on heuristic obfuscation. User P_1 sends privacy budget ε_1 to user P_2. P_2 compares ε_1 and ε_2. If $\varepsilon_1 > \varepsilon_2$, P_2 sends ε_2 to P_3. If $\varepsilon_1 < \varepsilon_2$, P_2 sends ε_1 to P_3. To reach better privacy-preserving, P_1 inspired P_2 sending smaller privacy budget to P_3.

- *Step* 3: By repeating the step 2, P_{n-1} and P_n negotiate a privacy budget ε_0 in a heuristic manner. P_n broadcasts the negotiating privacy budget ε_0 ($\varepsilon_0 \le \varepsilon_i$ for all users) to all users.

Under negotiating privacy budget ε_0, all users can obfuscate their own data by adding random noise. Thus, this process is called as heuristic obfuscation. Since the minimum privacy budget ε_0 is obtained by negotiating with all users, heuristic obfuscation can enhance privacy-preserving.

We give the algorithm 1 of negotiating privacy budget based on heuristic obfuscation.

Algorithm 1: Negotiating privacy budget based on heuristic obfuscation for mobile service based multiuser collaboration.

Input: Initializing privacy budget ε_i *for every user* P_i

Output: Negotiating privacy budget ε_0

```
Step 1: $epsilon = \varepsilon_1$
Step 2: for $i = 1 : n$
Step 3: if $(\varepsilon_i \le epsilon)$
Step 4: $epsilon = \varepsilon_i$
Step 5: end if
Step 6: end for
Step 7: $\varepsilon_0 = epsilon$
```

6.2. Framework Analysis

Now, we analyze the PPDAF. Since users add noise to their own data, server cannot obtain sensitive information of users by analyzing users' data or correlated data. Through negotiating, every user's expected privacy budget is at most negotiating privacy budget ε_0. Thus, the PPDAF guarantees every user's expected privacy budget, which is expected privacy-preserving level. According to the server's expected data utility, all users add noise to their own data ensuring expected data utility. In a word, the PPDAF can achieve Nash equilibrium between privacy and utility.

- *Theorem* 8 *(Nash Equilibrium of PPDAF)*. PPDAF maintains Nash equilibrium between expected data utility and negotiating privacy budget.
- *Proof*: Every user's expected privacy budget ε_i is at most ε_0 after completing the negotiating privacy budget based on heuristic obfuscation in PPDAF. Thus, every user P_i can achieve the negotiating privacy budget ε_0 at most. Every user P_i ensures privacy budget $\max\{\varepsilon_i\} \le \varepsilon_0$ based on Theorem 1, so every user P_i obtains expected privacy-preserving level. Since every user can get own noise dataset using adaptive Gaussian mechanism, approximate expected data utility AU is approximately equal to expected data utility U under conditional filtering noise $|Y| \in (0.5, 1.5)$. Thus, server can obtain approximate expected data utility AU of aggregation data and $AU \approx U$. So PPDAF maintains Nash equilibrium between approximate expected data utility U and negotiating privacy budget ε_0. Thus, (ε_0, AU) is a Nash equilibrium solution.

7. Experimental Evaluation

We make a comparative analysis by experimental evaluation in PPDAF. In all experimental evaluations, we initialize probability value $\delta = 0.01$ and sensitivity $\Delta f = 1$. All experiments are repeated 30 times for the validity of the results. We conduct these numerical experiments by implementing them with MATLAB (R2013b) and run our experiments on a desktop computer with Intel i5-2400 3.10 GHz processor, 4GB RAM, and Window 7 platform.

7.1. Dataset

We use T-Drive taxi trajectory dataset [27] to evaluate utility and privacy of PPDAF. This is a sample of T-Drive taxi trajectory dataset, which was generated by over 10,000 taxis in a period of one week in Beijing. We chose two datasets of Taxi ID 8 and Taxi ID 43 to conduct privacy and utility analysis. We evaluate the performance of PPDAF by using synthetic datasets of Taxi ID 568 and Taxi ID 569, Taxi ID 569 and Taxi ID 695, and Taxi ID 569 and Taxi ID 36. Every dataset is perturbed by using Gaussian mechanism, local differential privacy and adaptive Gaussian mechanism in all experiments. Then, we analyze privacy and utility of these perturbation datasets for PPDAF and evaluate the performance of PPDAF.

7.2. Privacy Analysis

We analyze the privacy-preserving level according to the expected estimation error. Because of directly perturbing data, the privacy metric is

$$EEE = \sum p(Y) \,|\, D' - D \,\|_1 = \sum p(Y) \,|\, Y \,| \qquad (9)$$

Where $D' = D + Y$. Since the trajectory data consist of latitude and longitude, the privacy metric is

$$EEE = \sum p(Y_1) p(Y_2) \,|\, Y_1 \,\|\, Y_2 \,| \qquad (10)$$

Where noise Y_1 and Y_2 added to the latitude and longitude, respectively.

From Figure 3, the expected estimation error curve using Gaussian mechanism decreases as privacy budget increasing to PPDAF. We can conclude that Gaussian mechanism may lead to utility disaster when privacy is very tiny. In PPDAF, when we use local differential privacy to achieve privacy preserving, we assume that the locations of trajectory data are true and code them as 1. Then, we achieve privacy preserving of trajectory datasets using local differential privacy with randomized response. The value of latitude of false response is generated randomly between the maximum latitude value and minimum latitude value in the all experimental trajectory datasets. Also, the value of longitude of false response is generated randomly between the maximum longitude value and minimum longitude value in the all experimental trajectory datasets. We can also observe that the expected estimated error decreases as privacy budget increases from Figure 4 and the expected estimated error is smaller. Thus, we can conclude that local differential privacy using randomized response leads to privacy leakage. Figure 5 shows that the expected estimation error curve using adaptive Gaussian mechanism also decreases as privacy budget increasing. The expected estimation error of PPDAF increases as the expected data utility decreases under the same privacy budget. The results show that PPDAF satisfies privacy-preserving monotonicity of differential privacy for different expected data utility. Thus, PPDAF achieves the expected privacy preserving under any expected data utility.

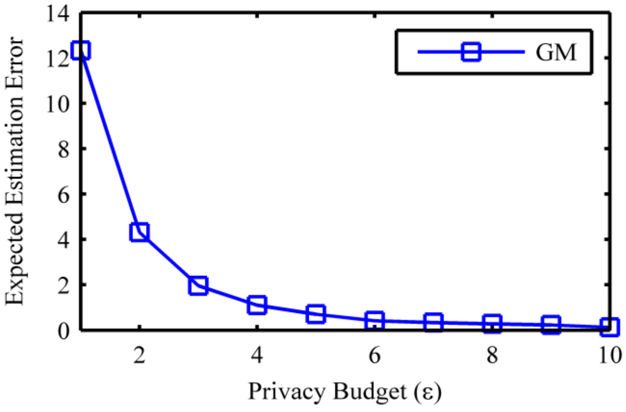

Figure 3. Expected estimation error curve using Gaussian mechanism (GM) for PPDAF.

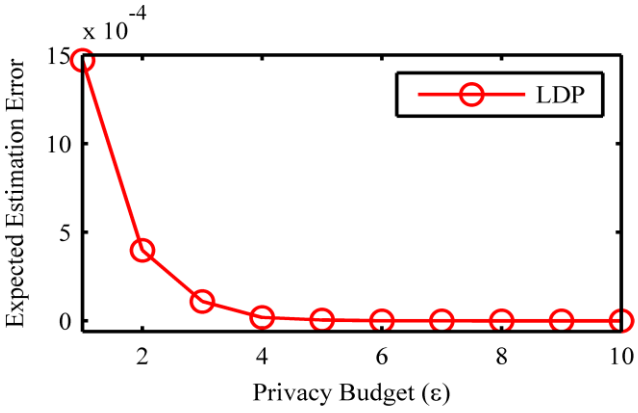

Figure 4. Expected estimation error curve using local differential privacy (LDP) for PPDAF.

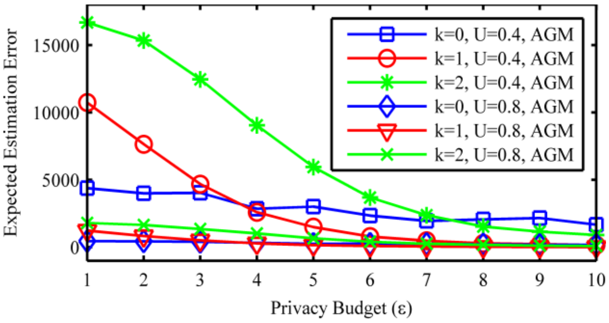

Figure 5. Expected estimation error curve using adaptive Gaussian mechanism (AGM) for PPDAF.

7.3. Utility Analysis

Now, we analyze the data utility of Gaussian mechanism and local differential privacy using randomized response based on utility metric $1-|x-x'|/|x|$ to the absolute value of relative error $(x-x')/x$, where x' is the random perturbed value of x. Also, we analyze the expected data utility of adaptive Gaussian mechanism for PPDAF according to Definition 6 of expected data utility metric.

Table 1. Data utility using Gaussian mechanism for PPDAF.

Trajectory Datasets	**Data Utility of Latitude and Longitude**	
	$\varepsilon=0.1$	$\varepsilon=1$
Taxi ID 8	(0.8507, 0.5349)	(0.9846, 0.9328)
	(0.8426, 0.4216)	(0.9855, 0.9408)
	(0.8501, 0.5326)	(0.9869, 0.9516)
	(0.8375, 0.5605)	(0.9860, 0.9536)
	(0.8591, 0.4744)	(0.9867, 0.9618)
Taxi ID 43	(0.8191, 0.5468)	(0.9792, 0.9601)
	(0.8636, 0.6256)	(0.9872, 0.9551)
	(0.8113, 0.4624)	(0.9855, 0.9656)
	(0.8069, 0.7240)	(0.9803, 0.9618)
	(0.8463, 0.5055)	(0.9897, 0.9565)

Table 2. Data utility using local differential privacy for PPDAF.

Trajectory Datasets	**Data Utility of Latitude and Longitude**	
	$\varepsilon=0.1$	$\varepsilon=1$
Taxi ID 8	(0.9613, 0.8908)	(0.9725, 0.9504)
	(0.9664, 0.9361)	(0.9694, 0.9194)
	(0.9755, 0.9291)	(0.9668, 0.9121)
	(0.9572, 0.9302)	(0.9560, 0.9167)
	(0.9589, 0.9368)	(0.9472, 0.8990)
Taxi ID 43	(0.9717, 0.9344)	(0.9637, 0.9518)
	(0.9654, 0.9537)	(0.9834, 0.9556)
	(0.9809, 0.9524)	(0.9823, 0.9618)
	(0.9739, 0.9299)	(0.9840, 0.9578)
	(0.9583, 0.9148)	(0.9649, 0.8912)

We observe data utility increases as privacy budget increases from Table 1 for PPDAF using Gaussian mechanism. When privacy budget is very small, the data utility of PPDAF using Gaussian mechanism is very bad result. For PPDAF using local differential privacy with randomized response, Table 2 shows that data utility is approximately equal to 1. Thus, local differential privacy using randomized response can achieve good data utility, but it could lead to privacy leakage and it cannot achieve expected data utility under any privacy budget. In Table 3, we observe that all approximate expected data utility approaches to expected data utility *U*=0.4 by using adaptive Gaussian mechanism for PPDAF. All approximate expected data utility approaches to expected data utility *U*=0.8 for PPDAF from Table 4. Thus, we can conclude that the approximate expected data utility is closer to the expected data utility as the absolute value *E* of relative error decreases. This result is consistent with Theorem 7. Thus, adaptive Gaussian mechanism can achieve approximate expected data utility under any expected privacy budget for PPDAF. We can conclude that adaptive Gaussian mechanism ensures Nash equilibrium between expected privacy budget and approximate expected data utility for PPDAF.

Furthermore, we compare the properties of Gaussian mechanism, local differential privacy, and adaptive Gaussian mechanism. In the Table 5, we observe that the adaptive Gaussian mechanism has better properties than Gaussian mechanism and local differential privacy. It not only satisfies differential privacy for any size dataset, but also maintains Nash equilibrium between privacy and utility. It can also prevent the third party making privacy attack of a distributed environment.

Table 3. Approximate expected data utility using adaptive Gaussian mechanism when expected data utility *U*=0.4 of PPDAF.

Trajectory Datasets	**Approximate Expected Data Utility of Latitude and Longitude**					
	$(k=0,\varepsilon=0.1)$	$(k=0,\varepsilon=1)$	$(k=1,\varepsilon=0.1)$	$(k=1,\varepsilon=1)$	$(k=2,\varepsilon=0.1)$	$(k=2,\varepsilon=1)$
Taxi ID 8	(0.3885, 0.4192)	(0.3956, 0.4416)	(0.3941, 0.3655)	(0.4289, 0.3855)	(0.4217, 0.3877)	(0.3689, 0.4528)
	(0.3502, 0.4343)	(0.4158, 0.4227)	(0.4018, 0.3591)	(0.4046, 0.4102)	(0.4240, 0.3915)	(0.4485, 0.3803)
	(0.4340, 0.4266)	(0.3456, 0.4066)	(0.4045, 0.3893)	(0.4154, 0.3665)	(0.4124, 0.4022)	(0.3905, 0.3504)
	(0.3815, 0.3763)	(0.3854, 0.4431)	(0.4441, 0.4224)	(0.4152, 0.4231)	(0.3988, 0.3835)	(0.3311, 0.3786)
	(0.3606, 0.3940)	(0.3376, 0.4217)	(0.4185, 0.4198)	(0.3911, 0.3540)	(0.4585, 0.3509)	(0.4118, 0.4082)
Taxi ID 43	(0.4157, 0.3931)	(0.4673, 0.4327)	(0.4222, 0.4000)	(0.3881, 0.3675)	(0.3680, 0.4229)	(0.3705, 0.3841)
	(0.4024, 0.4054)	(0.3796, 0.3989)	(0.3512, 0.4275)	(0.4621, 0.3996)	(0.4348, 0.3211)	(0.3862, 0.3927)
	(0.4628, 0.4156)	(0.4053, 0.4333)	(0.4020, 0.3884)	(0.4114, 0.4034)	(0.3580, 0.4435)	(0.3727, 0.3991)
	(0.4393, 0.3506)	(0.4375, 0.3698)	(0.4297, 0.4058)	(0.3879, 0.3802)	(0.4043, 0.3720)	(0.3801, 0.3595)
	(0.3838, 0.3744)	(0.3300, 0.4323)	(0.4534, 0.3929)	(0.4161, 0.4002)	(0.4020, 0.4169)	(0.3674, 0.3246)

Table 4. Approximate expected data utility using adaptive Gaussian mechanism when expected data utility *U*=0.8 of PPDAF.

Trajectory Datasets	**Approximate Expected Data Utility of Latitude and Longitude**					
	$(k=0,\varepsilon=0.1)$	$(k=0,\varepsilon=1)$	$(k=1,\varepsilon=0.1)$	$(k=1,\varepsilon=1)$	$(k=2,\varepsilon=0.1)$	$(k=2,\varepsilon=1)$
Taxi ID 8	(0.8014, 0.8170)	(0.8111, 0.8008)	(0.7994, 0.8068)	(0.8002, 0.7971)	(0.8132, 0.7883)	(0.7641, 0.7938)
	(0.7952, 0.8022)	(0.8073, 0.8063)	(0.8116, 0.8001)	(0.7974, 0.8036)	(0.7790, 0.7972)	(0.8043, 0.7995)
	(0.7927, 0.7954)	(0.7949, 0.8035)	(0.7815, 0.8018)	(0.7978, 0.8040)	(0.8109, 0.8101)	(0.7973, 0.7918)
	(0.7975, 0.8104)	(0.8001, 0.7977)	(0.8033, 0.7861)	(0.7856, 0.8120)	(0.7802, 0.8060)	(0.7890, 0.7859)
	(0.8076, 0.7820)	(0.7873, 0.7928)	(0.7888, 0.7955)	(0.8156, 0.8021)	(0.8014, 0.7990)	(0.7939, 0.8038)
Taxi ID 43	(0.8068, 0.7948)	(0.7972, 0.7962)	(0.8149, 0.8172)	(0.7954, 0.7974)	(0.7869, 0.8008)	(0.8006, 0.7749)
	(0.8074, 0.7985)	(0.8041, 0.8106)	(0.7900, 0.7983)	(0.7897, 0.7907)	(0.8004, 0.8157)	(0.7936, 0.7956)
	(0.8029, 0.8052)	(0.7921, 0.8075)	(0.8177, 0.7881)	(0.7934, 0.7928)	(0.8043, 0.7993)	(0.7955, 0.7926)
	(0.7895, 0.7879)	(0.8118, 0.8078)	(0.7869, 0.8048)	(0.7993, 0.8006)	(0.7914, 0.8140)	(0.7878, 0.7876)
	(0.7955, 0.7771)	(0.7812, 0.8088)	(0.8072, 0.7964)	(0.8100, 0.7942)	(0.7947, 0.8030)	(0.7915, 0.7937)

Table 5. Comparison between Gaussian mechanism, local differential privacy and adaptive Gaussian mechanism.

Mechanisms	**Dataset Size**	**Satisfying Differential Privacy**	**Maintaining Nash Equilibrium**	**Preventing Third Party from Privacy Attack**
Gaussian Mechanism	Any	Yes	No	Yes
Local Differential Privacy	Any	Yes	No	Yes
Adaptive Gaussian Mechanism	Any	Yes	Yes	Yes

7.4. Performance Analysis

We only consider the average running time cost of random perturbation to different trajectory datasets for performance evaluation of PPDAF using three differential privacy mechanisms. In Tables 6, 7, and 8, the PPDAF using Gaussian mechanism, local differential privacy and adaptive Gaussian mechanism keeps almost the same average running time cost under the same privacy budget and synthetic dataset size, respectively. But we can obtain that the average running time cost almost increases as privacy budget increases from Tables 6, 7, and 8. In Tables 7, and 8 we can conclude that the average running time cost is not affected by the expected data utility under the same privacy budget. Tables 6, 7, and 8 show that the average running time cost of PPDAF using adaptive Gaussian mechanism is the biggest. Since PPDAF using adaptive Gaussian mechanism can achieve expected data utility, this leads to need more time cost. Also, the average running time cost of PPDAF using adaptive Gaussian mechanism is almost decreases as $k \in [0,2]$ increases. Thus, PPDAF using adaptive Gaussian mechanism provides a tradeoff among expected privacy preserving, expected data utility and running time cost.

Table 6. Average running time cost of PPDAF using Gaussian mechanism and local differential privacy.

Synthetic Datasets	Dataset Size	Running Time (ms)			
		$\varepsilon = 0.1$		$\varepsilon = 1$	
		GM	LDP	GM	LDP
Taxi ID 568 and Taxi ID 569	362	0.1333	0.0110	0.1000	0.0110
Taxi ID 569 and Taxi ID 695	516	0.0667	0.0116	0.1333	0.0116
Taxi ID 569 and Taxi ID 36	675	0.0333	0.0119	0.1333	0.0119

Table 7. Average running time cost of PPDAF using adaptive Gaussian mechanism when privacy budget =0.1.

Synthetic Datasets	Dataset Size	Running Time (ms)					
		AGM, $\varepsilon = 0.1$, $U = 0.4$			AGM, $\varepsilon = 0.1$, $U = 0.8$		
		$k = 0$	$k = 1$	$k = 2$	$k = 0$	$k = 1$	$k = 2$
Taxi ID 568 and Taxi ID 569	362	144.7667	145.8667	139.2333	145.2333	147.2667	140.6333
Taxi ID 569 and Taxi ID 695	516	144.3333	146.0000	139.5667	145.4000	147.1667	140.6000
Taxi ID 569 and Taxi ID 36	675	145.5667	146.0667	139.3667	145.4000	147.1333	140.5000

Table 8. Average running time cost of PPDAF using adaptive Gaussian mechanism when privacy budget =1.

Synthetic Datasets	Dataset Size	Running Time (ms)					
		AGM, $\varepsilon = 1$, $U = 0.4$			AGM, $\varepsilon = 1$, $U = 0.8$		
		$k = 0$	$k = 1$	$k = 2$	$k = 0$	$k = 1$	$k = 2$
Taxi ID 568 and Taxi ID 569	362	232.7000	208.7000	159.6667	233.2000	208.7000	159.7333
Taxi ID 569 and Taxi ID 695	516	233.2667	208.7000	160.0667	233.4333	208.7333	159.7333
Taxi ID 569 and Taxi ID 36	675	233.6000	208.8333	160.1000	233.2667	208.5000	159.7333

8. Conclusions

Through our analysis, differential privacy and local differential privacy cannot satisfy Nash equilibrium between privacy and utility for mobile service based multiuser collaboration. Thus, we presented an adaptive Gaussian mechanism maintaining Nash equilibrium between privacy and utility. Then, we proposed PPDAF using adaptive Gaussian mechanism in a data aggregation environment. Also, we provide a method of negotiating privacy budget with heuristic obfuscation for PPDAF. Through theoretical analysis and experimental evaluation, the PPDAF can maintain Nash equilibrium between expected data utility and negotiating privacy budget. In adaptive Gaussian mechanism, the probability of satisfying conditional filtering noise is relatively small according to Theorem 5. Thus, the PPDAF using adaptive Gaussian mechanism requires high time cost for large scale dataset. Furthermore, the PPDAF using adaptive Gaussian mechanism is easy to extend to engineering implementation for a data aggregation environment.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 61602290 and 61662009), the Natural Science Basic Research Program of Shaanxi Province (No. 2017JQ6038), the Fundamental Research Founds for the Central Universities (No. GK201704016, 2016CBY004, GK201603093, GK261001236, and GK201501008), the Program of Key Science and Technology Innovation Team in Shaanxi Province (No. 2014KTC-18), and the Open Project Fund of Guizhou Provincial Key Laboratory of Public Big Data (No. 2017BDKFJJ026).

References

[1] Beimel A., Nissim K., and Omri E., "Distributed Private Data Analysis: Simultaneously Solving How and What," *in Proceedings of the International Cryptology Conference*, Santa Barbara, pp. 451-468, 2008.

[2] Chan T., Li M., Shi E., and Xu W., "Differentially Private Continual Monitoring of Heavy Hitters from Distributed Streams," *in Proceedings of the International Symposium on Privacy Enhancing Technologies Symposium*, Vigo, pp. 140-159, 2012.

[3] Cormode G., Jha S., Kulkarni T., Li N., Srivastava D., and Wang T., "Privacy at Scale: Local Differential Privacy in Practice," *in Proceedings of the ACM SIGMOD International Conference on Management of Data*, Houston, pp. 1655-1658, 2018.

[4] Duchi J., Jordan M., and Wainwright M., "Local Privacy and Statistical Minimax Rates," *in Proceeding of the 54th Annual IEEE Symposium on Foundations of Computer Science*, Berkeley, pp. 429-438, 2013.

[5] Dwork C., McSherry F., Nissim K., and Smith A., "Calibrating Noise to Sensitivity in Private Data Analysis," *in Proceedings of the Theory of Cryptography Conference*, New York, pp. 265-284, 2006.

[6] Dwork C., Kenthapadi K., McSherry F., Mironov I., and Naor M., "Our Data, Ourselves: Privacy via Distributed Noise Generation," *in Proceedings of the 24th Annual International Conference on the Theory and Applications of*

Cryptographic Techniques, Petersburg, pp. 486-503, 2006.

[7] Dwork C. and Roth A., "The Algorithmic Foundations of Differential Privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3-4, pp. 211-407, 2014.

[8] El-Sisi A., "Fast Cryptographic Privacy Preserving Association Rules Mining on Distributed Homogenous Database," *The International Arab Journal of Information Technology*, vol. 7, no. 2, pp. 152-160, 2010.

[9] Feild H., Allan J., and Glatt J., "Crowdlogging: Distributed, Private, and Anonymous Search Logging," *in Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Beijing, pp. 375-384, 2011.

[10] Hong Y., Vaidya J., Lu H., Karras P., and Goel S., "Collaborative Search Log Sanitization: Toward Differential Privacy and Boosted Utility," *IEEE Transactions on Dependable and Secure Computing*, vol. 12, no. 5, pp. 504-518, 2015.

[11] Jin X., Zhang R., Chen Y., Li T., and Zhang Y., "Dpsense: Differentially Private Crowdsourced Spectrum Sensing," *in Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, Vienna, pp. 296-307, 2016.

[12] Kasiviswanathan S., Lee H., Nissim K., Raskhodnikova S., and Smith A., "What Can We Learn Privately?" *SIAM Journal on Computing*, vol. 40, no. 3, pp. 793-826, 2011.

[13] Li C., Zhou P., Xiong L., Wang Q., and Wang T., "Differentially Private Distributed Online Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 8, pp. 1440-1453, 2018.

[14] Liu H., Wu Z., Peng C., Tian F., and Lu L., "Adaptive Gaussian Mechanism Based on Expected Data Utility under Conditional Filtering Noise," *KSII Transactions on Internet and Information Systems*, vol. 12, no. 7, pp. 3497-3515, 2018.

[15] McGregor A., Mironov I., Pitassi T., Reingold O., Talwar K., and Vadhan S., "The Limits of Two-Party Differential Privacy," *in Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science*, Las Vegas, pp. 81-90, 2010.

[16] McSherry F., "Privacy Integrated Queries: An Extensible Platform for Privacy-Preserving Data Analysis," *in Proceedings of the ACM SIGMOD International Conference on Management of Data*, Providence, pp. 19-30, 2009.

[17] McSherry F. and Mironov I., "Differentially Private Recommender Systems: Building Privacy into the Netflix Prize Contenders," *in Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, NW Washington, pp. 627-635, 2009.

[18] McSherry F. and Talwar K., "Mechanism Design Via Differential Privacy," *in Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, NW Washington, pp. 94-103, 2007.

[19] Mohammed N., Alhadidi D., Fung B., and Debbabi M., "Secure Two-Party Differentially Private Data Release for Vertically Partitioned Data," *Journal of the American Statistical Association*, vol. 11, no. 1, pp. 59-71, 2014.

[20] Sei Y. and Ohsuga A., "Differential Private Data Collection and Analysis Based on Randomized Multiple Dummies for Untrusted Mobile Crowdsensing," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 4, pp. 926-939, 2017.

[21] Sheffet O., "Locally Private Hypothesis Testing," *in Proceedings of the 35th International Conference on Machine Learning*, Stockholm, pp. 1-28, 2018.

[22] Shoham Y. and Leyton-Brown K., *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*, Cambridge University Press, 2008.

[23] Shokri R. and Shmatikov V., "Privacy-Preserving Deep Learning," *in Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, Denver, pp. 1310-1321, 2015.

[24] Smith A., Thakurta A., and Upadhyay J., "Is Interaction Necessary for Distributed Private Learning?" *in Proceedings of the IEEE Symposium on Security and Privacy*, San Jose, pp. 58-77, 2017.

[25] Su S., Tang P., Cheng X., Chen R and Wu Z., "Differentially Private Multi-Party High-Dimensional Data Publishing," *in Proceedings of the 32nd IEEE International Conference on Data Engineering*, Helsinki, pp. 205-216, 2016.

[26] Warner S., "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," *the American Statistical Association*, vol. 60, no. 309, pp. 63-69, 1965.

[27] Yuan J., Zheng Y., Xie X., and Sun G., "Driving with Knowledge From the Physical World," *in Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, pp. 316-324, 2011.

Hai Liu received his M.S. degree from Guizhou University, China, in 2015. Currently, he is a Ph.D. student in School of Computer Science, Shaanxi Normal University, China. His main research interest includes privacy protection.

Zhenqiang Wu received his Ph.D. degree from Xidian University, China, in 2007. He is currently a full professor of Shaanxi Normal University, China. His research interests include wireless networks, network security, and privacy protection.

Changgen Peng received his Ph.D. degree from Guizhou University, China, in 2007. He is currently a full professor of Guizhou University, China. His research interests include cryptography, information security, and privacy protection.

Feng Tian received his Ph.D. degree from Xi'an Jiaotong University, China, in 2015. He is currently a lecturer of Shaanxi Normal University, China. His research interest includes location privacy protection.

Laifeng Lu received her Ph.D. degree from Xidian University, China, in 2012. She is currently an associate professor of Shaanxi Normal University, China. Her research interests include privacy protection and network security.