This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2834232, IEEE Transactions on Information Forensics and Security

REAP: An Efficient Incentive Mechanism for Reconciling Aggregation Accuracy and Individual Privacy in Crowdsensing

Zhikun Zhang, *Student member, IEEE,* Shibo He, *Member, IEEE,*
Jiming Chen, *Senior Member, IEEE,* and Junshan Zhang, *Fellow, IEEE,*

***Abstract*—Incentive mechanism plays a critical role in privacy-aware crowdsensing. Most previous studies assume a trustworthy fusion center (FC) in their co-design of incentive mechanism and privacy preservation. Very recent work has taken step to relax the assumption on trustworthy FC and allowed participatory users (PUs) to randomly report their binary sensing data, whereas the focus is to examine PUs' equilibrium behavior. Making a paradigm shift, this paper aims to study the privacy compensation for continuous data sensing while allowing FC to directly control PUs. There are two conflicting objectives in such a scenario: FC desires better quality data in order to achieve higher aggregation accuracy whereas PUs prefer injecting larger noises for higher privacy-preserving levels (PPLs). To strike a good balance therein, we propose an efficient incentive mechanism named REAP to reconcile FC's aggregation accuracy and individual PU's data privacy. Specifically, we adopt the celebrated notion of differential privacy to quantify PUs' PPLs and characterize their impacts on FC's aggregation accuracy. Then, appealing to Contract Theory, we design an incentive mechanism to maximize FC's aggregation accuracy under a given budget. The proposed incentive mechanism offers different contracts to PUs with different privacy preferences, by which FC can directly control them. It can further overcome the *information asymmetry* problem, i.e., FC typically does not know each PU's precise privacy preference. We derive closed-form solutions for the optimal contracts in both *complete information* and *incomplete information* scenarios. Further, the results are generalized to the continuous case where PUs' privacy preferences take values in a continuous domain. Extensive simulations are provided to validate the feasibility and advantages of our proposed incentive mechanism.**

***Index Terms*—Crowd sensing, data aggregation, privacy preservation, incentive mechanism**

I. Introduction

THE recent proliferation of portable mobile devices (e.g., smartphone, smartwatch, tablet computer, etc.), integrated with a set of sensors (e.g., GPS, camera, accelerometer, etc.), has spurred much interest in mobile crowdsensing [1], [2]. Due to its advantage in reducing the deployment cost in large-scale sensing applications, crowdsensing has been applied to a large variety of areas such as smart transportation, environmental monitoring, health-care, etc [3]–[6].

Z. Zhang, S. He and J. Chen (Corresponding author) are with State Key Laboratory of Industrial Control Technology, Zhejiang University, and Cyber Innovation Joint Research Center, Hangzhou, China. E-mail: zhangzhk@zju.edu.cn, s18he@iipc.zju.edu.cn, cjm@zju.edu.cn

Junshan. Zhang is School of Electrical, Computer and Energy Engineering, Arizona State University, USA. E-mail: junshan.zhang@asu.edu

Typically, sensing data collected from participatory users (PUs) will be aggregated by the fusion center (FC) for data analytics. For example, to identify public health condition, FC can collect the daily exercise data from PUs and carry out data aggregation such as average and histogram. Clearly, contributing sensing data to FC can be costly for PUs, considering the resources being consumed (e.g., energy and bandwidth) and potential data privacy breach. Therefore, PUs would be reluctant to participate in crowdsensing without a proper incentive mechanism that compensates their cost. Most previous studies focused on resources consumption for data sensing and reporting in incentive mechanism design [7]–[9]. Only quite a few consider PUs' privacy losses [10], a common assumption made by these works is that FC is trustworthy such that PUs' privacy merely breach when FC releases the aggregation results to the public.

In practice, the assumption of trustworthy FC may not hold in many cases, e.g., when FC is compromised by malicious attackers, or the communication channels between PUs and FC are eavesdropped. Very recent work [11] has made the first attempt to relax the trustworthy FC assumption and study how to compensate PUs' privacy losses in a game-theoretic model. In [11], PUs can fully control their data privacy by adding well-calibrated noises to the raw sensing data before reporting them. However, the private data is assumed to be binary, which is not widely applicable to real-world system. Further, the focus of [11] is to examine the equilibrium behavior of PUs where FC has no direct control over them. Different from [11], this paper studies the privacy compensation for continuous data sensing, where FC has direct control over PUs' behaviors.

To this end, one challenge is to reconcile the following conflict: PUs prefer injecting larger noise for higher privacy preserving levels (PPLs) whereas FC desires better quality data for higher aggregation accuracy. Another challenge is to overcome the *information asymmetry* problem between FC and PUs, since it is difficult (perhaps impossible) to know PUs' privacy preferences. Further, privacy preferences of PUs are typically heterogenous, e.g., women have higher privacy preferences to their age than men, and patients are more concerned about their location privacy, which incurs diverse privacy losses for different PUs under the same PPL. An efficient incentive mechanism needs to differentiate the diverse privacy losses of PUs and provides appropriate rewards that capture their contributions to FC without knowing individual PU's precise privacy preference.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2834232, IEEE Transactions on Information Forensics and Security

To tackle these challenges, we propose REAP[1], an efficient incentive mechanism based on Contract Theory. By Contract Theory, FC can add some kind of enforcement to incentivize PUs by signing specific contracts with them. Different contracts are designed for PUs of different types, each of which specifies one type of PPL and the corresponding payment that a PU will receive if he/she can sacrifice the given PPL. A key concern here is to design a proper menu of contracts that satisfies incentive compatibility such that all PUs can maximize their utilities when they truthfully reveal their privacy preferences.

Specifically, we adopt differential privacy [12] to quantify individual privacy and (α, δ)-accuracy to measure FC's aggregation accuracy. Then, the quantitative impact of individual PU's PPL on FC's aggregation accuracy is derived. In light that each PU's impact on the aggregation accuracy is quantified, we can design a menu of optimal contracts that maximize FC's aggregation accuracy under a given budget. We first consider the *complete information* scenario where FC knows the precise type of each PU, and use the best achievable aggregation accuracy as the benchmark. We next consider the optimal contract design in *incomplete information* scenario where FC knows only the probability distribution of PUs' types. Closed-form solutions for both scenarios are derived. Further, we generalize our results to the case where PUs' privacy preferences can take value in a continuous domain. In such a case, the optimization problem turns out to be a functional extreme value problem that can be solved using an optimal control based approach.

The contributions of this paper are three folds:

1) We propose REAP, a Contract Theory based incentive mechanism, to compensate PUs' data privacy losses and hence resolve the information asymmetry issues between PUs and FC.
2) We adopt proper measures to quantify both individual PU's PPL and FC's aggregation accuracy, by which the quantitative impact of individual privacy on aggregation accuracy is derived.
3) Closed-form solutions are derived for both complete information and incomplete information scenarios. We also generalize our results to the case of continuous privacy preferences.

The rest of this paper is organized as follows. The related work is discussed in Section II. Section III presents an overview to the proposed crowdsensing system, and the quantitative impact of PUs' PPLs on FC's aggregation accuracy. In Section IV, we leverage Contract Theory to address the information asymmetry problem and generalize our results to the continuous case in Section V. Simulation results are illustrated in Section VI to validate our theoretical results. Section VII concludes this paper.

The main notations used in this paper are summarized in Table I.

[1]The name REAP comes from REconciling Aggregation accuracy and individual Privacy.

TABLE I: Notations

$\mathcal{U}$	Set of PUs
$\mathcal{D}$	Set of sensing data
d_i	PU i's raw sensing data
γ	PUs' data range
η_i	Laplacian noise added to PU i
b_i	Scale parameter of η_i
n	Number of PUs
k	Number of PUs' types
λ_k	Number of type-k PUs
u_i	PU i's utility
p_i	PU i's payment
ϵ_i	PU i's privacy preserving level
θ_i	PU i' privacy preference
α	FC's aggregation error
δ	FC's confidence level for the aggregation error
B	FC's budget constraint

II. Related Work

Recently, plenty of incentive mechanisms have been proposed to stimulate PUs' participation in mobile crowdsensing systems. Most of these mechanisms are based on either auction [10], [13]–[17] or other game-theoretic models [18]–[22], which aim to achieve different objectives. Specifically, in [15], [21], the authors aim to maximize the social welfare. The objective of [18], [19] is to maximize the profit of FC, and [17], [22] design mechanisms to minimize FC's payment. The basic requirement of these mechanisms is to guarantee that all PUs' costs are compensated, at least in the expectation sense. Most previous studies only compensate PUs' resources consumptions for sensing and reporting data, while their privacy losses are not remunerated explicitly.

Interestingly, Ghosh et al. took the first step to purchase PUs' privacy in their seminal work [23]. In [23], data owners bid their privacy preferences, and the system chooses a set of users and decides the corresponding PPLs to achieve the best statistic accuracy under a given budget. Based on this work, a few improved mechanisms [10], [24]–[26] have been proposed, especially considering the correlation between privacy preferences and private data. A common assumption made by these works is a trustworthy FC, where PUs should report their raw sensing data to FC. Differential privacy is employed to the aggregation result, i.e., the noise is added to the aggregation result. However, in our setting, we do not assume a trustworthy FC, and differential privacy is employed to each PU's raw sensing data, which is called local differential privacy. In practice, our setting is more practical, since FC may be compromised by malicious attackers, or the communication channels between PUs and FC may be eavesdropped. Furthermore, the mechanism design in our setting is more challenging, since the added noises will inevitably impair the aggregation accuracy. Therefore, the aggregation accuracy and FC's total payment should be jointly optimized instead of being designed separately as in previous works. Although [11] removed the trustworthy FC assumption, the focus of [11] is to examine PUs' equilibrium behavior, which may end up with an inefficient equilibrium, i.e., FC may not achieve desirable aggregation accuracy in a cost-efficient manner. Furthermore, the private data considered in [11] is binary bit, which is not widely applicable in mobile crowdsensing systems. Different

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2834232, IEEE Transactions on Information Forensics and Security

from [11], this paper aims to study the privacy compensation for continuous data sensing while allowing FC to directly control PUs' behavior, such that FC can achieve desirable aggregation accuracy in a cost-efficient manner.

Another line of related work is privacy-preserving mechanism design in mobile crowdsensing systems. These works do not take PUs' data privacy into consideration. Instead, they consider the privacy issue of the mechanism itself. For example, [27], [28] aim to preserve PUs' anonymity within the incentive mechanism, and [29]–[31] aims to preserve PUs' bidding privacy.

III. System Model

In this section, we first present the system overview. Then, we quantify PUs' PPLs and their impacts on FC's aggregation accuracy.

A. System Overview

The mobile crowdsensing system considered in this paper consists of an untrusted FC, a task agent and a set $\mathcal{U} = \{u_1, u_2, \cdots, u_n\}$ of PUs as shown in Fig. 1. FC aims to collect a set of sensing data from n PUs, denoted as $\mathcal{D} = \{d_1, d_2, \cdots, d_n\}$, where $d_i \in \mathbb{R}$ is a real number. Then it carries out some aggregation operations, such as average, max/min, histogram, etc, to abstract some valuable patterns. For easy exposition, we will investigate the average aggregation[2], i.e., $s = \frac{1}{n}\sum_{i=1}^{n} d_i$, which constitutes a large portion of currently deployed crowdsensing system. For example, some map application such as Baidu map collects GPS data (e.g., location and speed) from mobile vehicles and conducts average aggregation to estimate the real-time traffic condition. In the healthcare application, FC intends to collect PUs' daily exercise data and conduct average aggregation to monitor public health condition.

Clearly, the sensing data may contain sensitive information about PUs. A powerful and curious attacker may use these data to infer PUs' personal information. For example, in the healthcare application, the exercise data allow adversaries to infer individual PU's health condition or living habit. Different from most of the previous works on privacy-preserving data aggregation in crowdsensing, we do not assume FC to be trustworthy, since a sophisticated malicious attacker can compromise FC's database or eavesdrop the communication channels between PUs and FC. Therefore, PUs may not be willing to contribute their raw sensing data due to the privacy concern. To dispel PUs' worry about privacy, we propose to allow PUs to add well-calibrated noises η_i to their raw sensing data d_i before reporting them, and their PPLs can be strictly quantified by differential privacy as depicted in Section III-B.

However, there are two conflicting objectives in this setting: FC desires better quality data in order to achieve higher aggregation accuracy whereas PUs prefer adding larger noise for higher PPLs (these conflicts will further be quantified in Section III-C). In this paper, we aim to design an efficient mechanism to reconcile these conflicts. The framework of the proposed crowdsensing system is shown in Fig. 1 and the workflow is as follows:

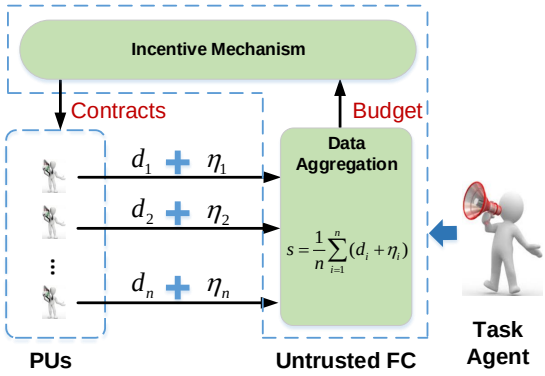


Fig. 1: Framework of REAP.

- Firstly, the task agent announces a sensing task to FC.
- **Incentive Mechanism.** Then, FC designs a menu of contract items (each specifies a privacy-payment pair) that maximize the aggregation accuracy under given budget, and broadcast them to all PUs. PUs can choose to sign any one of the contracts that maximize their own utilities. Once the contract is signed, PUs must report a privacy-preserving version of their sensing data with the PPLs specified in the contracts. In return, they will receive the corresponding payments.
- **Data Aggregation.** Next, upon receiving the privacy-preserving sensing data from PUs, FC conducts average aggregation on these data.
- Finally, FC returns the aggregated data to the task agent.

B. Differentially Private Data Reporting

In this subsection, we adopt the celebrated notion of differential privacy [32] to quantify PUs' PPLs and their corresponding privacy losses, with which we define PUs' utility function.

Informally, differential privacy guarantees that, after receiving the randomized data, the attackers cannot distinguish between two neighboring inputs with high confidence. Here, neighboring relationship is an important concept in differential privacy. In this paper, we adopt the neighboring relationship for continuous value as follows:

Definition 1 (γ_i-adjacency)**.** *Two continuous data d_i and d_i' are γ_i-adjacency, if $|d_i - d_i'| \leq \gamma_i$, where γ_i is the range of PU i's sensing data.*

Then, we can give the formal definition of differential privacy.

Definition 2 (ϵ_i-differential privacy [33])**.** *A random algorithm $\{\mathcal{A} : R \to R | \mathcal{A}(d_i) = d_i + \eta_i\}$ achieves ϵ_i-differential privacy, if for all pairs of γ_i-adjacency data d_i and d_i', and observation d^{obs},*

$$Pr[\mathcal{A}(d_i) = d^{obs}] \leq e^{\epsilon_i} Pr[\mathcal{A}(d_i') = d^{obs}]. \quad (1)$$

[2]We leave the discussion of other kinds of data aggregations in future work

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2834232, IEEE Transactions on Information Forensics and Security

Intuitively, PU i's accurate sensing data can be either d_i or d_i' from an attacker's view. As Fig. 2 shows, after adding noise η_i, both d_i and d_i' can result in d^{obs} with certain probabilities, i.e., p_i and p_i'. Thus, attackers cannot infer PU i's accurate sensing data with high confidence when they receive d^{obs}. Clearly, smaller ϵ_i means higher PPL, since it is more difficult to distinguish d_i and d_i' when observing d^{obs}.

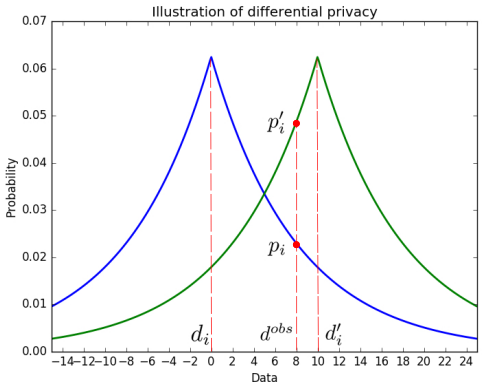


Fig. 2: Illustration of differential privacy.

Laplacian mechanism and exponential mechanism are most widely used mechanisms to achieve differential privacy [34, Chapter 2]. Laplacian mechanism is developed to handle the numeric queries while exponential mechanism is applied to non-numeric value queries, e.g., the output of query function is categorical. In this paper, the sensing data we considered are numeric data. Hence, we adopt the Laplacian mechanism to achieve differential privacy. The Laplacian mechanism achieves ϵ_i-differential privacy by calibrating the scale parameter following Lemma 1.

Lemma 1. *If the Laplacian mechanism is used, i.e., $\eta_i \sim Lap(0, b_i)$, we can achieve ϵ_i-differential privacy by setting $b_i = \frac{\gamma_i}{\epsilon_i}$.*

The proof can be found in Appendix A. By differential privacy, we can also define PUs' privacy loss. According to the utility theoretic characterization of differential privacy [35], the relationship between the expected utilities of two adjacent data can be characterized by e^{ϵ_i} based on (1). Since ϵ_i always takes a small value in practice, we have $e^{\epsilon_i} \approx 1 + \epsilon_i$. Further, the privacy loss can be modeled as the difference between the utility of true data and the utility of perturbed data [23], which is a linear function of ϵ_i according to the above observation. Then, we can define PUs' utility in Definition 3.

Definition 3 (PUs' utility)**.** *PU i's utility is given by*

$$u_i = p_i - \theta_i \epsilon_i, \tag{2}$$

where p_i is PU i's reward when he/she contributes sensing data to FC. θ_i is the privacy preference of PU i which indicates how much PUs care about their privacy. Clearly, different PUs have different privacy preferences [36], for instance, patients in hospital have higher privacy preferences to their location than others. Naturally, individual PU's privacy preference is private information and unknown to FC, or in other words, there exists *information asymmetry* between FC and PUs.

Notice that we only consider the cost incurred by PUs' privacy losses in order to ease the presentation of this paper, meanwhile the result in this paper can be adapted to incorporate other types of sensing costs. For instance, denote PU i's other types of costs by s_i. Then, his/her utility is given by $u_i = p_i - s_i - c_i$, where p_i and c_i is PU i's payment and privacy cost, respectively. Define $p_i' = p_i - s_i$. PU i's utility becomes $u_i = p_i' - c_i$, which is the same with the utility function in this paper.

C. Privacy versus Accuracy

In this subsection, we illustrate the conflict between FC's aggregation accuracy and PUs' PPLs by deriving their quantitative relationship.

To quantify the aggregation accuracy of the privacy preserving sensing data, we adopt the following accuracy definition.

Definition 4 ((α, δ)-accuracy)**.** *The aggregation $\hat{s}$ of privacy-preserving sensing data achieves (α, δ)-accuracy if*

$$Pr[|\hat{s} - s| \geq \alpha] \leq 1 - \delta,$$

where s is the aggregation result of accurate sensing data.

Intuitively, this definition indicates that the aggregation error is larger than α, with probability at most $1-\delta$. From estimation theory's perspective, α stands for confidence interval and δ stands for confidence level. Clearly, for a given confidence level, a smaller confidence interval / aggregation error means better aggregation accuracy. For ease of exposition, we leverage *aggregation error* α under a certain confidence level δ to represent the aggregation accuracy in the following part, where smaller aggregation error means better aggregation accuracy.

Then, we derive the quantitative relationship between individual PU's privacy and FC's aggregation accuracy as the following lemma:

Lemma 2. *For a given confidence level $\delta \leq 1$, the aggregation error α of the privacy-preserving sensing data is given by*

$$\alpha = \frac{\sqrt{2}\gamma}{n\sqrt{1-\delta}} \sqrt{\sum_{i=1}^{n} \frac{1}{\epsilon_i^2}}. \tag{3}$$

where ϵ_i is PU i's PPL, n is the number of PUs, and γ is the range of PUs' sensing data[3]. The proof can be found in Appendix B.

Recall that smaller ϵ_i indicates higher PPL. By examining Formula (3), we observe that FC's aggregation error α decreases when PUs adopt lower PPLs, i.e., larger ϵ_i, which conforms to our intuition. Formula (3) indicates that FC and PUs have conflicting objectives, i.e., FC hopes PUs to adopt lower PPLs to decrease the aggregation error, whereas PUs aim to adopt higher PPLs to better preserve their privacy. In the next section, we resolve this conflict through Contract Theory.

[3]Notice that the range of the sensing data should be the same for all PUs in a specific crowdsensing application, for example, the heart rate of a normal adult is always in the range $60 \sim 100$ bpm. Thus, all PUs' γ_i should take the same value, i.e., $\gamma_i = \gamma, \forall \gamma_i$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2834232, IEEE Transactions on Information Forensics and Security

IV. Incentive mechanism design: A Contract Theoretic Approach

So far, we have quantified the conflict between PUs' privacy and FC's aggregation accuracy. In this section, we present a contract theoretic approach to resolve the conflicting objectives between PUs and FC.

A. Contract Formulation

Contract theory generally studies how economic decision-makers construct contractual arrangement in the presence of *information asymmetry*. In this paper, FC typically does not know each PU's privacy preference θ_i, and aims to design a menu of contracts to stimulate PUs' participation in crowd-sensing. To facilitate later discussion, we classify PUs into different types based on their privacy preferences, i.e., the privacy preference of type-i PUs is θ_i.

In this section, we consider the case where PUs have finite types of privacy preferences, say k types $\Theta = \{\theta_i, \theta_2, \cdots \theta_k\}$. We leave the discussion of the case where θ takes continuous value in the next section. To facilitate the analysis, we sort PUs' types in ascending order, i.e., $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$. Using Contract theory, FC provides each type of PUs a given PPL ϵ_i and the corresponding payment p_i. Specifically, FC aims to design a set $\mathcal{C} = \{(\epsilon_1, p_1), \cdots, (\epsilon_k, p_k)\}$ of privacy-payment pairs called contract items. Each PU chooses to sign one of the contract items (ϵ_i, p_i), and then report ϵ_i-differentially private sensing data for payment p_i.

Each type of PUs choose the contract item that maximizes their utilities defined in (2). FC aims to optimize the contracts and minimize the aggregation error, i.e., minimize α derived in (3). Since $\frac{\sqrt{2}\gamma}{n\sqrt{1-\delta}}$ is a positive constant, minimizing $\alpha = \frac{\sqrt{2}\gamma}{n\sqrt{1-\delta}}\sqrt{\sum_{i=1}^{n}\frac{1}{\epsilon_i^2}}$ is equivalent to minimizing $\alpha = \sum_{i=1}^{n}\frac{1}{\epsilon_i^2}$.

In the following subsection, we will consider the optimal contract design under two information scenarios:

- **Complete information:** The complete information scenario is served as a benchmark, where FC knows each PU's precise type, and thus can offer a precise contract to each PU directly. Clearly, FC achieves lowest aggregation error in this scenario since there are no additional payments. This can serve as the lower bound of FC's achievable aggregation error in any information scenarios.
- **Incomplete information:** In the incomplete information scenario, FC does not know each PU's precise type, but knows the distribution of PUs' types, e.g., type-i has λ_i PUs. In this scenario, FC should design and broadcast a menu of optimal contracts to all PUs, and each PU can choose one of the contracts that maximize his/her utility.

B. Optimal Contract Design under Complete Information

In the complete information scenario, FC knows each PU's precise type. We will leverage the aggregation error achieved in this case as a benchmark to evaluate the performance of the proposed contracts under incomplete information scenario. As FC knows each PU's precise type, it can offer precise contract to each PU directly. In this scenario, FC only need to guarantee that each PU's utility is nonnegative so that they are willing to contribute their sensing data. In Contract Theory, we call this individual rationality constraints.

Definition 5 (Individual Rationality)**.** *A menu of contracts satisfy Individual Rationality (IR) constraints if they provide nonnegative utilities to all PUs, i.e.,*

$$p_i - \theta_i\epsilon_i \geq 0, \forall i. \tag{4}$$

Thus, we can design the optimal contracts under complete information by solving the following optimization problem:

Problem 1.

$$\min \sum_{i=1}^{k}\frac{\lambda_i}{\epsilon_i^2},$$
$$s.t. \sum_{i=1}^{k}\lambda_i p_i \leq B, \tag{5}$$
$$p_i - \theta_i\epsilon_i \geq 0, \quad \forall i. \tag{6}$$

where B is the total budget that FC possesses.

Then, we provide solution to this optimization problem.

Lemma 3. *The inequalities in (5) and (6) can take the equal sign simultaneously, i.e., $\sum_{i=1}^{k}\lambda_i p_i = B$ and $p_i - \theta_i\epsilon_i = 0$.*

It is easy to show that both (5) and (6) can take the equal sign by contradiction. Given p_i, if there exists a contract that satisfies $p_i - \theta_i\epsilon_i > 0$, then we can always find a larger ϵ_i to achieve lower aggregation error until the equality satisfies. Similarly, If there exists a contract that satisfies $\sum_{i=1}^{k}\lambda_i p_i < B$, we can always find a larger p_i, which means larger ϵ_i, to achieve lower aggregation error until the equality satisfies, which leads to the correctness of this lemma.

Lemma 3 shows that both IR constraints and budget constraint are tight at the optimal solution to Problem (1), which indicates that FC can provide zero utility to each type-i PU with $p_i^* = \theta_i\epsilon_i^*$ and expend the budget. Therefore, Problem 1 can be reduced to the following problem:

Problem 2.

$$\min \sum_{i=1}^{k}\frac{\lambda_i}{\epsilon_i^2},$$
$$s.t. \sum_{i=1}^{k}\lambda_i p_i = B, \tag{7}$$
$$p_i - \theta_i\epsilon_i = 0, \quad \forall i. \tag{8}$$

By solving Problem 2, we have the following theorem.

Theorem 4. *In the complete information scenario, the optimal contract $\{\epsilon_i^*, p_i^*\}$ is given by*

$$\epsilon_i^* = \frac{B}{\sum_{j=1}^{k}\lambda_j\theta_j^{\frac{2}{3}}}\theta_i^{-\frac{1}{3}}, \tag{9}$$
$$p_i^* = \frac{B}{\sum_{j=1}^{k}\lambda_j\theta_j^{\frac{2}{3}}}\theta_i^{\frac{2}{3}}. \tag{10}$$

The proof can be found in Appendix C. By looking into the parameters in the optimal contracts provided in Theorem 4, we have the following observation.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2834232, IEEE Transactions on Information Forensics and Security

Observation 1. *Recall that smaller ϵ_i means higher PPL, Theorem 4 illustrates that type-i PU's PPL decreases with B, and increases with θ_i, which conforms to our intuition. In another word, more budgets stimulate PUs to choose lower PPLs to achieve lower aggregation error, and FC tends to buy less privacy from PUs with higher privacy preferences to reduce payment.*

C. Optimal Contract Design under Incomplete Information

In the incomplete information scenario, FC does not know each PU's precise type, while the distribution of PUs' types is assumed to be known, i.e., type-i have λ_i PUs. In practice, the distribution of PUs' types can be obtained through questionnaire survey or analysis of the historical behavior of PUs [37], [38]. Clearly, FC should design an optimal contract for each type of PUs to achieve lower aggregation error, but due to the lack of knowledge about each PU's precise type, FC can only broadcast all contracts to all PUs. However, if choosing the contract designed for other types of PUs brings them higher utilities, some selfish PUs may pretend to be other types. To encourage all PUs to truthfully reveal their types, the optimal contracts should guarantee that choosing the contract corresponding to their own type can always achieve the highest utilities. Formally, we define this requirement as incentive compatibility constraints.

Definition 6 (Incentive Compatibility)**.** *A menu of contracts satisfies Incentive Compatibility (IC) constraints if the contract designed for type-i PUs brings them highest utility, i.e.,*

$$p_i - \theta_i\epsilon_i \geq p_j - \theta_i\epsilon_j, \quad \forall j \neq i. \tag{11}$$

Apart from the incentive compatibility constraints, the contracts under incomplete information scenario should also satisfy the individual rationality constraints in Definition 5. Thus, we can design the optimal contract under incomplete information scenario by solving the following optimization problem:

Problem 3.

$$\begin{aligned} \min \ & \sum_{i=1}^{k} \frac{\lambda_i}{\epsilon_i^2}, \\ s.t. \ & \sum_{i=1}^{k} \lambda_i p_i \leq B, && (12) \\ & p_i - \theta_i\epsilon_i \geq 0, \quad \forall i, && (13) \\ & p_i - \theta_i\epsilon_i \geq p_j - \theta_i\epsilon_j, \quad \forall j \neq i. && (14) \end{aligned}$$

Notice that in Problem 3, there are k IR constraints and $k(k-1)$ IC constraints, which makes it difficult to solve the optimization problem. Next, we show that these constraints can be reduced to a set of fewer equivalent constraints by the following lemmas.

Lemma 5. *The k IR constraints can be reduced to the following one constraint:*

$$p_k - \theta_k\epsilon_k = 0. \tag{15}$$

Proof. Recall that we have sorted PUs' types in ascending order, i.e., $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$. Based on IC constraints, we have

$$p_i - \theta_i\epsilon_i \geq p_k - \theta_i\epsilon_k \geq p_k - \theta_k\epsilon_k, \forall i \neq k.$$

Thus, if the IR constraint of type-k satisfied, i.e., $p_k - \theta_k\epsilon_k \geq 0$, it will satisfied for all other types automatically. Therefore, we can keep the last IR constraint and reduce the others. Moreover, if there exists an optimal contract that satisfies $p_k - \theta_k\epsilon_k > 0$, we can always find a larger ϵ_k to achieve lower aggregation error until $p_k - \theta_k\epsilon_k = 0$, which concludes the proof. □

Lemma 5 shows that only the highest type of PUs achieve zero utilities, and lower types of PUs achieve positive utilities that decrease with their types. The reason is that FC does not know each PU's type, it needs to provide incentives in terms of positive utilities to PUs to attract them revealing their truthful types. This is called *information loss* compared to complete information.

Lemma 6 (Monotonic Property)**.** *If $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$, then $\epsilon_1 \geq \epsilon_2 \geq \cdots \geq \epsilon_k$ holds.*

Proof. By the IC constraints, it is directly that

$$\begin{aligned} p_i - \theta_i\epsilon_i &\geq p_j - \theta_i\epsilon_j, \\ p_j - \theta_j\epsilon_j &\geq p_i - \theta_j\epsilon_i. \end{aligned}$$

Adding these two inequalities, we have $\epsilon_i(\theta_j - \theta_i) \geq \epsilon_j(\theta_j - \theta_i)$. Thus, if $\theta_i \leq \theta_j$, then $\epsilon_i \geq \epsilon_j$ for all i and j, which leads to the correctness of this lemma. □

Intuitively, Lemma 6 illustrates that PUs with higher type should be assigned lower PPL, since their unit cost is higher and FC needs to compensate them more when their impacts on the aggregation accuracy are the same. Further, this Lemma can be leveraged to prove the correctness of Lemma 7.

Lemma 7. *The $k(k-1)$ IC constraints can be reduced to the following $k-1$ constraints.*

$$p_i - \theta_i\epsilon_i = p_{i+1} - \theta_i\epsilon_{i+1}, \forall i \leq k-1. \tag{16}$$

The proof can be found in Appendix D. Lemma 7 ensures that if the contract item (ϵ_i, p_i) designed for type-i PUs bring them the same utilities with the contract item $(\epsilon_{i+1}, p_{i+1})$ designed for type-$(i+1)$ PUs, all the IC constraints for type-i PUs are satisfied.

Based on Lemma 5 and Lemma 7, we can reduce Problem 3 to the following problem:

Problem 4.

$$\begin{aligned} \min \ & \sum_{i=1}^{k} \frac{\lambda_i}{\epsilon_i^2}, \\ s.t. \ & \sum_{i=1}^{k} \lambda_i p_i = B, && (17) \\ & p_k - \theta_k\epsilon_k = 0, && (18) \\ & p_i - \theta_i\epsilon_i = p_{i+1} - \theta_i\epsilon_{i+1}, \forall i \leq k-1. && (19) \end{aligned}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2834232, IEEE Transactions on Information Forensics and Security

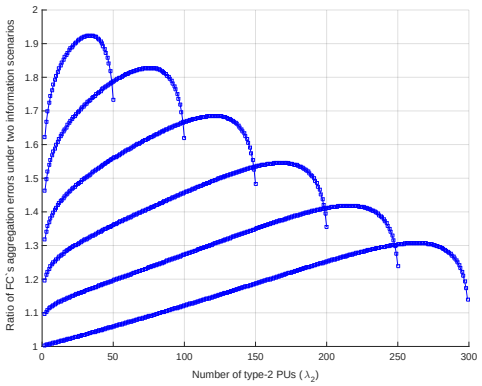


Fig. 3: The ratio of FC's lower aggregation error under incomplete information and complete information as a function of PUs' realization in three types, i.e., $\frac{\alpha_I}{\alpha_C}$.

By solving Problem 4, we can calculate the optimal contracts with the following theorem.

Theorem 8. *In the incomplete information scenario, the optimal contract* $\{\epsilon_i^*, p_i^*\}$ *is given by*

$$\epsilon_i^* = G H_i^{-\frac{1}{3}} \lambda_i^{\frac{1}{3}},$$

$$p_i^* = \begin{cases} G(\theta_i H_i^{-\frac{1}{3}} \lambda_i^{\frac{1}{3}} + \sum_{j=i+1}^{k} \Delta\theta_j H_j^{-\frac{1}{3}} \lambda_j^{\frac{1}{3}}), & i \neq k, \\ G\theta_k H_k^{-\frac{1}{3}} \lambda_k^{\frac{1}{3}}, & i = k, \end{cases}$$

where

$$\Delta\theta_i = \theta_i - \theta_{i-1}, \tag{20}$$

$$H_i = \begin{cases} \lambda_1\theta_1, & i = 1, \\ \lambda_i\theta_i + \Delta\theta_i \sum_{j=1}^{i-1} \lambda_j, & i > 1, \end{cases} \tag{21}$$

$$G = \frac{B}{\sum_{j=1}^{k} H_j^{\frac{2}{3}} \lambda_j^{\frac{1}{3}}}. \tag{22}$$

The proof of Theorem 8 is given in Appendix E. Next, we compare FC's aggregation accuracy under incomplete and complete information scenarios. In Fig. 3, we show the ratio of FC's aggregation accuracy under incomplete information and complete information scenarios when there are three types. $\lambda_1 = 0, 50, 100, 150, 200, 250$ correspond to the lines from bottom to top, respectively. In this figure, we only show λ_1 and λ_2, and $\lambda_3 = N - \lambda_1 - \lambda_2$. Other parameters are $N = 300, B = 1000, \gamma = 10, \delta = 0.9, \theta_1 = 1, \theta_2 = 2, \theta_3 = 3$. The ratio is a function of PUs' realization $\{\lambda_i\}_{i=1}^3$ in three types, which is always larger than or equal to 1, as FC achieves lower aggregation error under complete information scenario. By analyzing Fig. 3, we have the following observation.

Observation 2. *Compared with complete information, FC achieves higher aggregation error under incomplete information. The gap between FC's aggregation error under two information scenarios is minimized when all PUs belong to the highest type, i.e., type-3. For fixed* λ_1*, the gap increases when the number of type-3 PUs decrease until they reach a small value.*

The ratio reaches 1 when all PUs belong to the highest type, since in this situation, all PUs obtain zero utilities as in the complete information scenario. When the number of type-3 PUs decreases, the information loss increases, which leads to the increase of the gap. However, when the number of type-3 PUs reaches a small value, the effect of information loss decreases compared to the complete information, so that the ratio increases.

D. Discussions on Practical Implementation

By solving the above optimization problems, FC could provide a menu of optimal contracts to stimulate all types of PUs' participation in crowdsensing. However, PUs' action, if cannot be monitored by FC, may deviate from the contract in practice, e.g., a selfish PU may add noise with higher PPL than which signed in the contract to achieve higher utility. To ensure that all PUs generate noise strictly with the PPLs signed in the contract, FC can install a trusted crowdsensing app in PUs' mobile devices [39]. Once the contract is signed, the noise level would be controlled by the trusted crowdsensing app, such that PUs' PPLs can be monitored by FC.

Unlike traditional crowdsensing systems that aim to select a subset of PUs to conduct a specific task (e.g., reporting whether the traffic is jam or not), this paper focuses on the data aggregation, where FC aims to collect enough sensing data from PUs to conduct statistic analysis. Specifically, we investigate the average aggregation of a group of specified PUs' sensing data (e.g., calculating the average time of exercises in a specific area). From a statistical perspective, we aim to study the calculation of the population mean of the specified PUs' sensing data. Therefore, we need to collect all PUs' sensing data to yield the true population mean. In practice, if the population of PUs is extremely large, we can sample a subset of PUs to estimate the population mean. By [40], we know that the average of the sample is an unbiased estimation of the population mean. However, we emphasize that unbiased estimation is not equivalent to the true population mean. If the population of PUs is not large enough, it is recommended to involve all PUs' participation to calculate the true population mean.

V. Generalization to the Continuous Case

In this section, we analyze the optimal contracts design when PUs' privacy preferences take values from continuous domain.

We assume that PUs' types θ are in the interval $[\underline{\theta}, \overline{\theta}]$, and the probability density function of θ is $h(\theta)$. Similar to the analysis of discrete case, FC can design the optimal contracts by solving the following optimization problem:

Problem 5.

$$\min \int_{\underline{\theta}}^{\overline{\theta}} \frac{h(\theta)}{\epsilon^2(\theta)} d\theta,$$

$$s.t. \int_{\underline{\theta}}^{\overline{\theta}} p(\theta)h(\theta)d\theta \leq B, \tag{23}$$

$$p(\theta) - \theta\epsilon(\theta) \geq 0, \tag{24}$$

$$p(\theta) - \theta\epsilon(\theta) \geq p(\hat{\theta}) - \theta\epsilon(\hat{\theta}), \forall\hat{\theta} \neq \theta. \tag{25}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2834232, IEEE Transactions on Information Forensics and Security

where (23) is the budget constraint, (24) is the IR constraints and (25) is the IC constraints.

Notice that the IR and IC constraints in (24) and (25) are infinite since θ is a continuous value, making the optimization problem challenging. Similarly, we reduce the IR and IC constraints by the following two lemmas.

Lemma 9. *The infinite IR constraints can be reduced to the following one constraint,*

$$p(\overline{\theta}) - \overline{\theta}\epsilon(\overline{\theta}) = 0. \tag{26}$$

Proof. By the the IC constraints, we can derive the following inequality,

$$\begin{aligned} p(\theta) - \theta\epsilon(\theta) &\geq p(\overline{\theta}) - \theta\epsilon(\overline{\theta}) \\ &\geq p(\overline{\theta}) - \overline{\theta}\epsilon(\overline{\theta}), \forall \theta \neq \overline{\theta}. \end{aligned}$$

The above inequality shows that the IR constraint of type-$\overline{\theta}$ satisfied implies the IR constraints satisfied for all $\theta \in [\underline{\theta}, \overline{\theta}]$. Thus, we can reduce IR constraints to $p(\overline{\theta}) - \overline{\theta}\epsilon(\overline{\theta}) \geq 0$. Further, we show that $p(\overline{\theta}) - \overline{\theta}\epsilon(\overline{\theta}) \geq 0$ can take the equal sign. If there exists a contract $(\epsilon(\overline{\theta}), p(\overline{\theta}))$ such that $p(\overline{\theta}) - \overline{\theta}\epsilon(\overline{\theta}) > 0$, we can always find a larger $\epsilon(\overline{\theta})$ to achieve lower aggregation error until $p(\overline{\theta}) - \overline{\theta}\epsilon(\overline{\theta}) = 0$, which leads to correctness of this lemma. □

Lemma 10. *The infinite IC constraints can be reduced to the following two constraints,*

$$\frac{d\epsilon(\theta)}{d\theta} \leq 0, \tag{27}$$

$$\frac{dp(\theta)}{d\theta} - \theta\frac{d\epsilon(\theta)}{d\theta} = 0. \tag{28}$$

Proof. Based on (25), we can derive the following two local conditions for type-θ PUs,

$$\frac{dp(\hat{\theta})}{d\hat{\theta}}\Big|_{\hat{\theta}=\theta} - \theta\frac{d\epsilon(\hat{\theta})}{d\hat{\theta}}\Big|_{\hat{\theta}=\theta} = 0, \tag{29}$$

$$\frac{d^2p(\hat{\theta})}{d\hat{\theta}^2}\Big|_{\hat{\theta}=\theta} - \theta\frac{d^2\epsilon(\hat{\theta})}{d\hat{\theta}^2}\Big|_{\hat{\theta}=\theta} \leq 0. \tag{30}$$

Since (29)(30) hold for all $\theta \in [\underline{\theta}, \overline{\theta}]$, we have

$$\frac{dp(\theta)}{d\theta} - \theta\frac{d\epsilon(\theta)}{d\theta} = 0, \tag{31}$$

$$\frac{d^2p(\theta)}{d\theta^2} - \theta\frac{d^2\epsilon(\theta)}{d\theta^2} \leq 0. \tag{32}$$

By differentiating (31), we can simplify (32) as

$$\frac{d\epsilon(\theta)}{d\theta} \leq 0. \tag{33}$$

Then, we prove that (31) and (33) hold globally. By integrating (31) from $\hat{\theta}$ to θ, we have

$$p(\theta) - p(\hat{\theta}) = \theta\epsilon(\theta) - \theta\epsilon(\hat{\theta}) - \int_{\hat{\theta}}^{\theta} \epsilon(u)du. \tag{34}$$

Rearrange (34), we have

$$p(\theta) - \theta\epsilon(\theta) = p(\hat{\theta}) - \hat{\theta}\epsilon(\hat{\theta}) + (\hat{\theta} - \theta)\epsilon(\hat{\theta}) - \int_{\hat{\theta}}^{\theta} \epsilon(u)du.$$

Since $\epsilon(\theta)$ is non-increasing, we have $(\hat{\theta} - \theta)\epsilon(\hat{\theta}) - \int_{\hat{\theta}}^{\theta} \epsilon(u)du \geq 0$. Thus, we can conclude that $p(\theta) - \theta\epsilon(\theta) \geq p(\hat{\theta}) - \hat{\theta}\epsilon(\hat{\theta})$ for all $\hat{\theta} \neq \theta$, which indicates that (31) and (33) hold globally. □

Similar to the analysis of the discrete case, the budget constraint (23) can take the equal sign, i.e.,

$$\int_{\underline{\theta}}^{\overline{\theta}} p(\theta)h(\theta)d\theta = B. \tag{35}$$

Then, we can transform Problem 5 to the following problem:

Problem 6.

$$\begin{aligned} \min \ & \int_{\underline{\theta}}^{\overline{\theta}} \frac{h(\theta)}{\epsilon^2(\theta)}d\theta, \\ s.t. \ & (35)(26)(27)(28). \end{aligned}$$

Notice that Problem 6 is a functional extreme value problem, we can utilize the optimal control method to solve this problem.

Let $u(\theta) = \epsilon(\theta)$ be the control variable, and let $x_1(\theta) = p(\theta) - \theta\epsilon(\theta)$ be the state variable. Then, we have

$$\begin{aligned} \dot{x}_1(\theta) &= \dot{p}(\theta) - \epsilon(\theta) - \theta\dot{\epsilon}(\theta) \\ &= -\epsilon(\theta) = -u(\theta), \end{aligned}$$

where the second equality is due to (28).

To deal with the budget constraint (23), we can define a new state variable

$$\dot{x}_2(\theta) = p(\theta)h(\theta) = [x_1(\theta) + \theta u(\theta)]h(\theta) \tag{36}$$

Based on (23), we can derive the following transversality condition,

$$x_2(\overline{\theta}) - x_2(\underline{\theta}) = B. \tag{37}$$

Thus, the Hamiltonian of the optimal control problem is given by

$$\begin{aligned} &H[x(\theta), u(\theta), \lambda(\theta), \theta] \\ =&\frac{h(\theta)}{u^2(\theta)} - \lambda_1(\theta)u(\theta) + \lambda_2(\theta)[x_1(\theta) + \theta u(\theta)]h(\theta), \end{aligned}$$

where $\lambda_1(\theta)$ and $\lambda_2(\theta)$ are co-state variables.

According to the Euler-Lagrange equation for optimal control problem, we have the following conditions,

$$\begin{aligned} \frac{\partial H}{\partial u} &= \frac{-2h(\theta)}{u^3(\theta)} - \lambda_1 + \lambda_2\theta h(\theta) = 0, \\ \dot{\lambda}_1(\theta) &= -\frac{\partial H}{\partial x_1} = -\lambda_2 h(\theta), \\ \dot{\lambda}_2(\theta) &= -\frac{\partial H}{\partial x_2} = 0. \end{aligned}$$

Thus, we can calculate the co-state variables as,

$$\begin{aligned} \lambda_2(\theta) &= c_1, \\ \lambda_1(\theta) &= -c_1 H(\theta) + c_2, \end{aligned}$$

where c_1 and c_2 are constants which can be calculated by the transversality conditions (37) and (26).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2834232, IEEE Transactions on Information Forensics and Security

Then, the optimal contract $[\epsilon^*(\theta), p^*(\theta)]$ is given by,

$$\begin{aligned}\epsilon^*(\theta) &= u^*(\theta) \\ &= \sqrt[3]{\frac{2h(\theta)}{c_1\theta h(\theta) - c_1 H(\theta) - c_2}}, \\ p^*(\theta) &= x_1(\theta) + \theta\epsilon(\theta) \\ &= \theta\epsilon^*(\theta) - \int_{\underline{\theta}}^{\theta} \epsilon^*(\tau)d\tau.\end{aligned}$$

VI. Simulation Studies

In this section, we first validate the feasibility of the proposed contracts, and then analyze the impact of different system parameters on the aggregation error.

TABLE II: Simulation settings

Parameter	Value	
Number of PUs (n)	200	
Privacy preference (θ)	$[5, 15]$	
Number of PUs' types (k)	Feasibility	20
	Performance	$[5, 20]$
Budget constraint (B)	Feasibility	1000
	Performance	$[500, 1000]$

The simulation settings are shown in Table II. We assume there are 200 PUs and their privacy preferences are from 5 to 15. For simplicity, we assume PUs' privacy preferences follow uniform distribution.

To illustrate the feasibility of the proposed contracts, we show that the proposed optimal contracts satisfy both *monotonic* property and *incentive compatibility* property, which are discussed in Lemma 6 and Definition 6, respectively. The number of PUs' types k and the budget constraint B are set to 20 and 1000, respectively. To evaluate the impact of parameter k and B on the aggregation error (defined in Lemma 2), we set their value ranges to $[5, 20]$ and $[500, 1000]$, respectively.

A. Contract Feasibility

In this subsection, we illustrate that the proposed optimal contracts satisfy both *monotonic* property and *incentive compatibility* property.

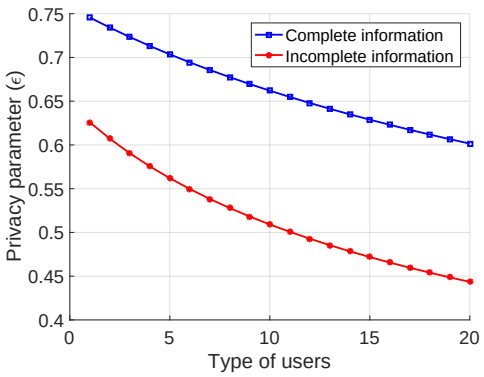


Fig. 4: Contract monotonicity.

Fig. 4 shows that ϵ decreases when PUs' types increase. Since smaller ϵ means higher PPL, Fig. 4 indicates that PUs with higher types tend to choose higher PPLs, which validates the *monotonic* property. Besides, the result conforms to our intuition that FC chooses to purchase less privacy from PUs with higher privacy preferences to reduce the payment. On another hand, we find that under the same budget constraint, PUs' PPLs under complete information scenario are lower than which in incomplete information scenario. The reason is that in complete information scenario, FC knows each PU's precise type, such that the contract designed for all types of PUs can take zero utilities, as Lemma 3 shows. However, in the incomplete information scenario, PUs' precise types are unknown to FC. Thus, only the highest type PUs achieve zero utilities, whereas other types of PUs' utilities remain strictly positive, since otherwise, lower type PUs will pretend to be higher types to achieve higher utilities.

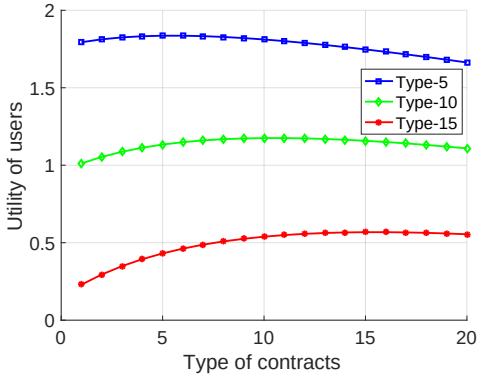


Fig. 5: Contract incentive compatibility.

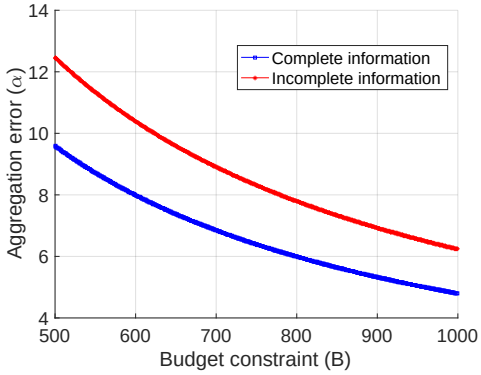


Fig. 6: Aggregation accuracy Vs. budget constraint.

In Fig. 5, we show the utility function of type-5, type-10 and type-15 PUs when they choose different types of contracts. Notice that the utility functions are concave for all types of PUs, and each type of PUs achieve their optimal utilities when they choose their corresponding contract, e.g., type-5 PUs achieve their optimal utilities when they choose type-5 contract, which validates the *incentive compatibility* property. Additionally, we observe that PUs with lower type achieve higher utilities when they choose the same contract. The reason is that the lower type PUs have lower privacy preference θ_i, according to PUs' utility definition $u_j = p_j - \theta_i\epsilon_j, \forall j$, smaller θ_i result in higher utility.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2834232, IEEE Transactions on Information Forensics and Security

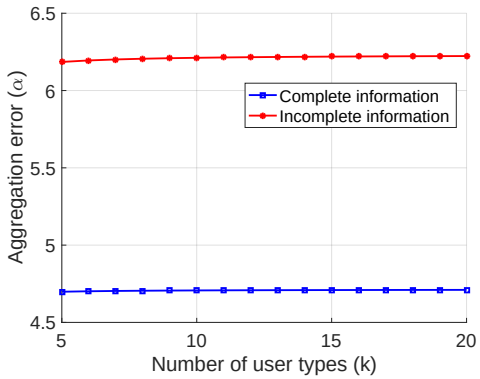


Fig. 7: Aggregation accuracy Vs. number of user types.

B. System Performance

In this subsection, we show the impacts of different system parameters on the aggregation error.

Fig. 6 shows the impact of the amount of budget on the aggregation error when other parameters are fixed. We observe that the aggregation error α decreases when the amount of budget increases, which indicates that larger amount of budget leads to lower aggregation error. The reason is obvious, when FC possesses more budget, it can provide more incentive to stimulate PUs to choose lower PPLs, leading to lower aggregation error.

In Fig. 7, we evaluate the impact of the number of PUs' types on the aggregation error when other parameters are fixed. Fig. 7 shows that, the aggregation error decreases with the number of PUs' types. By the reduced IR constraint $p_k - \theta_k \epsilon_k = 0$ and IC constraints $p_i - \theta_i \epsilon_i = p_{i+1} - \theta_i \epsilon_{i+1}$, FC can set the utilities of higher types of PUs more close to 0, meaning less additional payments. That is to say, the increase of PUs' types results in more additional payments, which decreases the aggregation error under a given budget.

VII. Conclusion

In this paper, we designed an incentive mechanism named REAP to compensate PUs' privacy losses. Unlike previous mechanisms, we did not require FC to be trustworthy and allow PUs to add well calibrated noises to their sensing data before reporting them. Then, in order to achieve better aggregation accuracy under a budget constraint, we devised a contract-based incentive mechanism to induce PUs to choose lower PPLs. Optimal contracts with closed form were derived in both complete and incomplete information scenarios. Our results were generalized to the continuous case. Extensive simulations were conducted to validate the feasibility and performance of our proposed incentive mechanism.

Acknowledgments

This work is supported by NSFC under Grant 61429301, U1401253.

References

[1] X. Duan, C. Zhao, S. He, P. Cheng, and J. Zhang, "Distributed algorithms to compute walrasian equilibrium in mobile crowdsensing," *IEEE Transactions on Industrial Electronics*, 2016.

[2] S. He, D.-H. Shin, J. Zhang, and J. Chen, "Near-optimal allocation algorithms for location-dependent tasks in crowdsensing," *IEEE Transactions on Vehicular Technology*, 2016.

[3] R. K. Ganti, N. Pham, H. Ahmadi, S. Nangia, and T. F. Abdelzaher, "Greengps: a participatory sensing fuel-efficient maps application," in *Proceedings of ACM MobiSys'10*, pp. 151–164.

[4] J. Chen, K. Hu, Q. Wang, Y. Sun, Z. Shi, and S. He, "Narrowband internet of things: Implementations and applications," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 2309–2314, 2017.

[5] Y. Cheng, X. Li, Z. Li, S. Jiang, Y. Li, J. Jia, and X. Jiang, "Aircloud: a cloud-based air-quality monitoring system for everyone," in *Proceedings of ACM SenSys'14*, pp. 251–265.

[6] S. Hu, L. Su, H. Liu, H. Wang, and T. F. Abdelzaher, "Smartroad: Smartphone-based crowd sensing for traffic regulator detection and identification," *ACM Transactions on Sensor Networks*, vol. 11, no. 4, p. 55, 2015.

[7] G. Yang, S. He, Z. Shi, and J. Chen, "Promoting cooperation by the social incentive mechanism in mobile crowdsensing," *IEEE Communications Magazine*, vol. 55, no. 3, pp. 86–92, 2017.

[8] Q. Zhang, Y. Wen, X. Tian, X. Gan, and X. Wang, "Incentivize crowd labeling under budget constraint," in *Proceedings of IEEE INFOCOM'15*, pp. 2812–2820.

[9] X. Zhang, G. Xue, R. Yu, D. Yang, and J. Tang, "Truthful incentive mechanisms for crowdsourcing," in *Proceedings of IEEE INFOCOM'15*, pp. 2830–2838.

[10] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, "Inception: incentivizing privacy-preserving data aggregation for mobile crowd sensing systems," in *Proceedings of ACM MobiHoc'17*, pp. 341–350.

[11] W. Wang, L. Ying, and J. Zhang, "The value of privacy: Strategic data subjects, incentive mechanisms and fundamental limits," pp. 249–260.

[12] X. Yang, T. Wang, X. Ren, and W. Yu, "Survey on improving data utility in differentially private sequential data publishing," *IEEE Transactions on Big Data*, 2017.

[13] D. Yang, G. Xue, X. Fang, and J. Tang, "Incentive mechanisms for crowdsensing: Crowdsourcing with smartphones," *IEEE/ACM Transactions on Networking*, 2015.

[14] M. Zhang, L. Yang, X. Gong, and J. Zhang, "Privacy-preserving crowdsensing: Privacy valuation, network effect, and profit maximization," in *Proceedings of IEEE GLOBECOM'16*, pp. 1–6.

[15] H. Jin, L. Su, D. Chen, K. Nahrstedt, and J. Xu, "Quality of information aware incentive mechanisms for mobile crowd sensing systems," in *Proceedings of ACM MobiHoc'15*, pp. 167–176.

[16] X. Zhang, Z. Yang, Z. Zhou, H. Cai, L. Chen, and X. Li, "Free market of crowdsourcing: Incentive mechanism design for mobile sensing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 12, pp. 3190–3200, 2014.

[17] I. Koutsopoulos, "Optimal incentive-driven design of participatory sensing systems," in *Proceedings of IEEE INFOCOM'13*, pp. 1402–1410.

[18] L. Duan, T. Kubo, K. Sugiyama, J. Huang, T. Hasegawa, and J. Walrand, "Incentive mechanisms for smartphone collaboration in data acquisition and distributed computing," in *Proceedings of IEEE INFOCOM'12*, pp. 1701–1709.

[19] T. Luo, S. S. Kanhere, H.-P. Tan, F. Wu, and H. Wu, "Crowdsourcing with tullock contests: A new perspective," in *Proceedings of IEEE INFOCOM'15*, pp. 2515–2523.

[20] D. Peng, F. Wu, and G. Chen, "Pay as how well you do: A quality based incentive mechanism for crowdsensing," in *Proceedings of ACM MobiHoc'15*, pp. 177–186.

[21] M. H. Cheung, R. Southwell, F. Hou, and J. Huang, "Distributed time-sensitive task selection in mobile crowdsensing," in *Proceedings of ACM MobiHoc'15*, pp. 157–166.

[22] H. Xie, J. Lui, W. Jiang, and W. Chen, "Incentive mechanism and protocol design for crowdsensing systems," in *Allerton*, 2014.

[23] A. Ghosh and A. Roth, "Selling privacy at auction," in *Proceedings of ACM EC'11*, pp. 199–208.

[24] L. K. Fleischer and Y.-H. Lyu, "Approximately optimal auctions for selling privacy when costs are correlated with data," in *Proceedings of ACM EC'12*, pp. 568–585.

[25] K. Ligett and A. Roth, "Take it or leave it: Running a survey when privacy comes at a cost," in *Proceedings of WINE'12*. Springer, pp. 378–391.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2834232, IEEE Transactions on Information Forensics and Security

[26] K. Nissim, S. Vadhan, and D. Xiao, "Redrawing the boundaries on purchasing data from privacy-sensitive individuals," in *Proceedings of ACM ITCS'14*, pp. 411–422.
[27] Q. Li and G. Cao, "Providing efficient privacy-aware incentives for mobile sensing," in *Proceedings of IEEE ICDCS'14*, pp. 208–217.
[28] ——, "Providing privacy-aware incentives for mobile sensing," in *Proceedings of IEEE PerCom'13*, pp. 76–84.
[29] H. Jin, L. Su, B. Ding, K. Nahrstedt, and N. Borisov, "Enabling privacy-preserving incentives for mobile crowd sensing systems," in *Proceedings of IEEE ICDCS'16*, pp. 344–353.
[30] R. Zhu and K. G. Shin, "Differentially private and strategy-proof spectrum auction with approximate revenue maximization," in *Proceedings of IEEE INFOCOM'15*, 2015, pp. 918–926.
[31] R. Zhu, Z. Li, F. Wu, K. Shin, and G. Chen, "Differentially private spectrum auction with approximate revenue maximization," in *Proceedings of ACM MobiHoc'14*, pp. 185–194.
[32] C. Dwork, "Differential privacy: A survey of results," in *Proceedings of TAMC'08*. Springer, pp. 1–19.
[33] J. Le Ny and G. J. Pappas, "Differentially private filtering," *IEEE Transactions on Automatic Control*, vol. 59, no. 2, pp. 341–354, 2014.
[34] N. Li, M. Lyu, D. Su, and W. Yang, "Differential privacy: From theory to practice," *Synthesis Lectures on Information Security, Privacy, & Trust*, vol. 8, no. 4, pp. 1–138, 2016.
[35] M. M. Pai and A. Roth, "Privacy and mechanism design," *ACM SIGecom Exchanges*, vol. 12, no. 1, pp. 8–29, 2013.
[36] L. Xu, C. Jiang, Y. Chen, Y. Ren, and K. R. Liu, "Privacy or utility in data collection? a contract theoretic approach," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 7, pp. 1256–1269, 2015.
[37] Y. Zhang, L. Song, W. Saad, Z. Dawy, and Z. Han, "Contract-based incentive mechanisms for device-to-device communications in cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2144–2155, 2015.
[38] L. Duan, L. Gao, and J. Huang, "Cooperative spectrum sharing: a contract-based approach," *IEEE Transactions on Mobile Computing*, vol. 13, no. 1, pp. 174–187, 2014.
[39] H. Zhuo, S. Zhong, and N. Yu, "A privacy-preserving remote data integrity checking protocol with data dynamics and public verifiability," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 9, pp. 1432–1437, 2011.
[40] S. K. Sengijpta, "Fundamentals of statistical signal processing: Estimation theory," 1995.

APPENDIX A
PROOF OF LEMMA 1

Proof. Since the Laplacian mechanism is adopted,

$$\frac{Pr[\mathcal{A}(d_i)=d^{obs}]}{Pr[\mathcal{A}(d_i')=d^{obs}]}=\frac{f(d^{obs};d_i,b_i)}{f(d^{obs};d_i',b_i)},$$

where $f(\cdot;\mu,\sigma)$ is the probability density function of the Laplacian random variables with mean μ and variance $2\sigma^2$.

By the definition of γ_i-adjacency, we have,

$$\frac{f(d^{obs};d_i,b_i)}{f(d^{obs};d_i',b_i)}\leq e^{\frac{\gamma_i}{b_i}},$$

By the definition of differential privacy, it is clear that $\frac{\gamma_i}{b_i}=\epsilon_i$, which concludes the proof. □

APPENDIX B
PROOF OF LEMMA 2

Proof. The aggregation error of the randomized sensing data can be expressed as

$$\hat{s}-s=\frac{1}{n}\sum_{i=1}^{N}(d_i+\eta_i)-\frac{1}{n}\sum_{i=1}^{N}d_i=\frac{1}{n}\sum_{i=1}^{N}\eta_i.$$

Recall that the variance of Laplacian random variable $\eta_i\sim Lap(0,b_i)$ is $2b_i^2$, i.e., $D(\eta_i)=2b_i^2$, we can derive that

$$D(\frac{1}{n}\sum_{i=1}^{N}\eta_i)=\frac{2}{n^2}\sum_{i=1}^{n}b_i^2.$$

Therefore, from the *Chebyshev's inequality*, we have

$$P[|s-\hat{s}|\geq\alpha]\leq\frac{2}{\alpha^2n^2}\sum_{i=1}^{n}b_i^2,$$

which indicates that the aggregated randomized sensing data satisfies $(\alpha,\frac{2}{\alpha^2n^2}\sum_{i=1}^{n}b_i^2)$-accuracy.

Thus, for a given confidence level $\delta\leq 1$, we have

$$\alpha=\frac{\sqrt{2}\gamma}{n\sqrt{1-\delta}}\sqrt{\sum_{i=1}^{n}b_i^2}$$

Substituting $b_i=\frac{\gamma_i}{\epsilon_i}$ into the above formula, and set $\gamma_i=\gamma$ for all i, we have

$$\alpha=\frac{\sqrt{2}\gamma}{n\sqrt{1-\delta}}\sqrt{\sum_{i=1}^{n}\frac{1}{\epsilon_i^2}}.$$

□

APPENDIX C
PROOF OF THEOREM 4

Proof. Substituting (8) to (7), we have

$$\sum_{i=1}^{k}\lambda_i\theta_i\epsilon_i=B \tag{38}$$

The Lagrangian of Problem 2 is given by

$$L(\epsilon_i,\alpha)=\sum_{i=1}^{k}[\frac{\lambda_i}{\epsilon_i^2}+\alpha\lambda_i\theta_i\epsilon_i]-\alpha B,$$

where α is the Lagrangian multiplier.

Based on the KKT condition, we have

$$\frac{\partial L}{\partial\epsilon_i}=\frac{-2\lambda_i}{\epsilon_i^3}+\alpha\lambda_i\theta_i=0,\quad\forall i.$$

Solving the above equation obtain $\epsilon_i=\sqrt[3]{\frac{2}{\alpha}}\theta_i^{-\frac{1}{3}}$. Substituting this formula to (38), we have

$$\sqrt[3]{\frac{2}{\alpha}}=\frac{B}{\sum_{i=1}^{k}\lambda_i\theta_i^{\frac{2}{3}}}.$$

Therefore, ϵ_i^* is given by

$$\epsilon_i^*=\frac{B}{\sum_{j=1}^{k}\lambda_j\theta_j^{\frac{2}{3}}}\theta_i^{-\frac{1}{3}}, \tag{39}$$

Substituting (39) to $p_i^*-\theta_i\epsilon_i^*=0$, p_i^* can be calculated as

$$p_i^*=\frac{B}{\sum_{j=1}^{k}\lambda_j\theta_j^{\frac{2}{3}}}\theta_i^{\frac{2}{3}}. \tag{40}$$

□

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2834232, IEEE Transactions on Information Forensics and Security

APPENDIX D
PROOF OF LEMMA 7

Proof. We conduct the proof of this lemma by three steps.

Firstly, we prove that if $p_i - \theta_i\epsilon_i \geq p_{i-1} - \theta_i\epsilon_{i-1}$ satisfies, then $p_i - \theta_i\epsilon_i \geq p_j - \theta_i\epsilon_j$ hold for all $j \in \{i-1, i-2, \cdots, 1\}$.

Based on the IC constraints, we have

$$p_i - \theta_i\epsilon_i \geq p_{i-1} - \theta_i\epsilon_{i-1}, \tag{41}$$

$$p_{i-1} - \theta_{i-1}\epsilon_{i-1} \geq p_{i-2} - \theta_{i-1}\epsilon_{i-2}. \tag{42}$$

Formula (42) can be transformed to the following form

$$\theta_{i-1}(\epsilon_{i-2} - \epsilon_{i-1}) \geq p_{i-2} - p_{i-1}.$$

Recall the monotonic property in Lemma 6, we know that $\theta_{i-1} \leq \theta_i$ and $\epsilon_{i-2} \geq \epsilon_{i-1}$. Thus, we have $\theta_i(\epsilon_{i-2} - \epsilon_{i-1}) \geq p_{i-2} - p_{i-1}$ or $p_{i-1} - \theta_i\epsilon_{i-1} \geq p_{i-2} - \theta_i\epsilon_{i-2}$. Following the same step, we have

$$p_i - \theta_i\epsilon_i \geq p_{i-1} - \theta_i\epsilon_{i-1} \geq \cdots \geq p_1 - \theta_i\epsilon_1.$$

These inequalities lead to the correctness of this step.

Secondly, we prove that if $p_i - \theta_i\epsilon_i \geq p_{i+1} - \theta_i\epsilon_{i+1}$ satisfies, then $p_i - \theta_i\epsilon_i \geq p_j - \theta_i\epsilon_j$ hold for all $j \in \{i+1, i+2, \cdots, k\}$.

Similar to the proof of the first step, we have

$$p_i - \theta_i\epsilon_i \geq p_{i+1} - \theta_i\epsilon_{i+1} \geq \cdots \geq p_1 - \theta_i\epsilon_1,$$

which leads to the correctness of this step. Notice that for an optimal contract, we have $p_i - \theta_i\epsilon_i = p_{i+1} - \theta_i\epsilon_{i+1}$ holds, since otherwise, we can always find a larger ϵ_i to achieve lower aggregation error until the equal signs hold.

Thirdly, we prove that $p_i - \theta_i\epsilon_i = p_{i+1} - \theta_i\epsilon_{i+1}$ implies $p_i - \theta_i\epsilon_i \geq p_{i-1} - \theta_i\epsilon_{i-1}$.

It is obvious that $\theta_i(\epsilon_{i-1} - \epsilon_i) \geq \theta_{i-1}(\epsilon_{i-1} - \epsilon_i)$, rearrange this inequality, we have

$$p_i - \theta_i\epsilon_i \geq p_i + \theta_{i-1}\epsilon_{i-1} - \theta_{i-1}\epsilon_i - \theta_i\epsilon_{i-1}.$$

Since $p_i - \theta_i\epsilon_i = p_{i+1} - \theta_i\epsilon_{i+1}$, then $p_{i-1} - \theta_{i-1}\epsilon_{i-1} = p_i - \theta_{i-1}\epsilon_i$ hold, i.e., $p_i + \theta_{i-1}\epsilon_{i-1} - \theta_{i-1}\epsilon_i = p_{i-1}$. Thus, we have $p_i - \theta_i\epsilon_i \geq p_{i-1} - \theta_i\epsilon_{i-1}$.

In summary, $p_i - \theta_i\epsilon_i = p_{i+1} - \theta_i\epsilon_{i+1}$ implies $p_i - \theta_i\epsilon_i \geq p_j - \theta_i\epsilon_j, \forall j \neq i$, which ends the proof of this lemma. □

APPENDIX E
PROOF OF THEOREM 8

Proof. Based on (18) and (19), we have

$$\begin{aligned} p_{k-1} - \theta_{k-1}\epsilon_{k-1} &= p_k - \theta_{k-1}\epsilon_k \\ &= \theta_k\epsilon_k - \theta_{k-1}\epsilon_k \\ &= (\theta_k - \theta_{k-1})\epsilon_k \end{aligned} \tag{43}$$

Let $\Delta\theta_k = \theta_k - \theta_{k-1}$, we can rewrite (43) as $p_{k-1} = \theta_{k-1}\epsilon_{k-1} + \Delta\theta_k\epsilon_k$.

Following the same procedure, we can conclude that

$$p_i = \begin{cases} \theta_i\epsilon_i + \sum_{j=i+1}^{k} \Delta\theta_j\epsilon_j, & i \neq k, \\ \theta_k\epsilon_k, & i = k, \end{cases} \tag{44}$$

where $\Delta\theta_i$ is defined by (20).

Then, we have

$$\begin{aligned} \sum_{i=1}^{k} \lambda_i p_i =& \sum_{i=1}^{k-1}[\lambda_i\theta_i\epsilon_i + \lambda_i \sum_{j=i+1}^{k} \Delta\theta_j\epsilon_j] + \lambda_k\theta_k\epsilon_k \\ =& \lambda_k\theta_k\epsilon_k + \lambda_{k-1}\theta_{k-1}\epsilon_{k-1} + \lambda_{k-1}\Delta\theta_k\epsilon_k \\ &+ \lambda_{k-2}\theta_{k-2}\epsilon_{k-2} + \lambda_{k-2}[\Delta\theta_{k-1}\epsilon_{k-1} + \Delta\theta_k\epsilon_k] \\ &\vdots \\ &+ \lambda_1\theta_1\epsilon_1 + \lambda_1[\Delta\theta_2\epsilon_2 + \cdots + \Delta\theta_k\epsilon_k] \\ =& \epsilon_k[\lambda_k\theta_k + \Delta\theta_k(\lambda_{k-1} + \cdots + \lambda_1)] \\ &+ \epsilon_{k-1}[\lambda_{k-1}\theta_{k-1} + \Delta\theta_{k-1}(\lambda_{k-2} + \cdots + \lambda_1)] \\ &\vdots \\ &+ \epsilon_1\lambda_1\theta_1. \end{aligned}$$

Rearrange the above equation by ϵ_i, we can get

$$\sum_{i=1}^{k} \lambda_i p_i = \sum_{i=1}^{k} H_i\epsilon_i = B, \tag{45}$$

where H_i is defined by (21).

Thus, the Lagrangian of Problem 4 is

$$L(\epsilon, \alpha) = \sum_{i=1}^{k}[\frac{\lambda_i}{\epsilon_i^2} + \alpha H_i\epsilon_i] - \alpha B,$$

where α is the Lagrangian multiplier.

Based on the KKT condition, we have

$$\frac{\partial L}{\partial \epsilon_i} = \frac{-2\lambda_i}{\epsilon_i^3} + \alpha H_i = 0.$$

Then, we can calculate ϵ_i as

$$\epsilon_i = \sqrt[3]{\frac{2}{\alpha}}(\frac{\lambda_i}{H_i})^{\frac{1}{3}} \tag{46}$$

Substituting (46) to (45), we obtain

$$\sqrt[3]{\frac{2}{\alpha}} = \frac{B}{\sum_{j=1}^{k} H_j^{\frac{2}{3}}\lambda_j^{\frac{1}{3}}}$$

Thus, the optimal contract ϵ_i^* is given by

$$\epsilon_i^* = \frac{B}{\sum_{i=1}^{k} H_i^{\frac{2}{3}}\lambda_i^{\frac{1}{3}}} H_i^{-\frac{1}{3}}\lambda_i^{\frac{1}{3}} \tag{47}$$

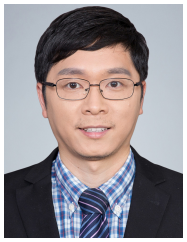
Then, we can calculate the k-th contract as,

$$p_k^* = \theta_k\epsilon_k^* = \frac{B}{\sum_{j=1}^{k} H_j^{\frac{2}{3}}\lambda_j^{\frac{1}{3}}}\theta_k H_k^{-\frac{1}{3}}\lambda_k^{\frac{1}{3}}.$$

Substitute (47) to (44) and rearrange, we can achieve other contracts when $i \neq k$. □

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2834232, IEEE Transactions on Information Forensics and Security

Zhikun Zhang received the B.Eng. degree in automation in 2014 from Shandong University, Jinan, China. From Oct. 2017 to Oct. 2018, he is a visiting scholar with Purdue University, West Lafayette, IN, USA. He is currently working toward the Ph.D. degree in the Group of Networked Sensing and Control (IIPC-NeSC) in the State Key Laboratory of Industrial Control Technology, Zhejiang University. His research interests include mechanism design, differential privacy and its applications in cognitive radio, crowdsensing system and machine learning.



Shibo He (M'13) received the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2012. From Nov. 2010 to Nov. 2011, he was a visiting scholar with the University of Waterloo, Waterloo, ON, Canada. He was an Associate Research Scientist from March 2014 to May 2014, and a postdoctoral scholar from May 2012 to February 2014, with Arizona State University, Tempe, AZ, USA. He is currently a Professor at Zhejiang University. His research interests include wireless sensor networks, crowdsensing and big data analysis.

Dr. He serves on the editorial board of IEEE Transactions on Vehicular Technology, Springer Peer-to-Peer Networking and Application, KSII transactions Internet and Information Systems, and is a guest editor of Elsevier Computer Communications and Hindawi International Journal of Distributed Sensor Networks. He served as publicity chair for IEEE SECON 2016, Registration and Fiance chair for ACM MobiHoc 2015, TPC Co-chair for IEEE ScalCom 2014, TPC Vice Co-chair for ANT 2013-2014, Track Co-chair for the Pervasive Algorithms, Protocols, and Networks of EUSPN 2013, Web Co-Chair for IEEE MASS 2013, and Publicity Co-chair of IEEE WiSARN 2010. Dr. He is the recipient of IEEE Asia-Pacific outstanding researcher award, 2015.

Jiming Chen (M'08-SM'11) received B.Sc. degree and Ph.D. degree both in Control Science and Engineering from Zhejiang University in 2000 and 2005, respectively. He was a visiting researcher at University of Waterloo from 2008 to 2010. Currently, he is a Changjiang Scholars Chair Professor (MOE) with College of Control Science and Engineering, deputy director of State Key Laboratory of Industrial Control Technology, and member of academic committee at Zhejiang University, China. He serves/served associate editors for several international Journals including IEEE Transactions on Parallel and Distributed System, IEEE Network, IEEE Transactions on Control of Network Systems, etc. He is the recipient of Fok Ying Tung Young Teacher Award of Ministry of Education, IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award, etc. He is IEEE VTS Distinguished Lecturer. His research interests include Internet of things, sensor networks, networked control, control system security, etc.

Junshan Zhang (Fellow) received his Ph.D. degree from the School of ECE at Purdue University in 2000. He joined the School of ECEE at Arizona State University in August 2000, where he has been Fulton Chair Professor since 2015. His research interests fall in the general field of information networks and data science, including communication networks, Internet of Things (IoT), Fog Computing, social networks, smart grid. His current research focuses on fundamental problems in information networks and data science, including Fog Computing and its applications in IoT and 5G, IoT data privacy/security, optimization/control of mobile social networks, cognitive radio networks, stochastic modeling and control for smart grid.

Prof. Zhang is a Fellow of the IEEE, and a recipient of the ONR Young Investigator Award in 2005 and the NSF CAREER award in 2003. He received the IEEE Wireless Communication Technical Committee Recognition Award in 2016. His papers have won a few awards, including the Kenneth C. Sevcik Outstanding Student Paper Award of ACM SIGMETRICS/IFIP Performance 2016, the Best Paper Runner-up Award of IEEE INFOCOM 2009 and IEEE INFOCOM 2014, and the Best Paper Award at IEEE ICC 2008 and ICC 2017. Building on his research findings, he co-founded Smartiply Inc in 2015, a Fog Computing startup company delivering boosted network connectivity and embedded artificial intelligence for IoT applications.

Prof. Zhang was TPC co-chair for a number of major conferences in communication networks, including IEEE INFOCOM 2012 and ACM MOBIHOC 2015. He was the general chair for ACM/IEEE SEC 2017, WiOPT 2016, and IEEE Communication Theory Workshop 2007. He was a Distinguished Lecturer of the IEEE Communications Society. He was an Associate Editor for IEEE Transactions on Wireless Communications, an editor for the Computer Network journal, and an editor IEEE Wireless Communication Magazine. He is currently serving as an editor-at-large for IEEE/ACM Transactions on Networking and an editor for IEEE Network Magazine.