A Game-Theoretic Approach to Quality Control for Collecting Privacy-Preserving Data

Weina Wang, Lei Ying, and Junshan Zhang
School of Electrical, Computer and Energy Engineering
Arizona State University, Tempe, AZ 85287
{weina.wang, lei.ying.2, junshan.zhang}@asu.edu

***Abstract*—We consider the design of an incentive mechanism for a principal, *who is assumed to be not trustworthy*, to collect informative data from privacy-sensitive individuals. The principal offers payments to incentivize participation and informative data reporting. The individuals are strategic and take into account both the payment and the cost for privacy loss during data reporting. Due to privacy concerns, an individual may be willing to report only a "noisy" version of the private data, resulting in quality degradation. To achieve desirable accuracy of data analysis, it is imperative for the principal to have an incentive mechanism under which the quality of the collected data is controllable.**

In this paper, we exploit a game-theoretic approach to the design of a payment mechanism, such that the quality of the collected data is controllable through a parameter ϵ by making sure that each individual's strategy in a Nash equilibrium is to participate and symmetrically randomize her data, while guaranteeing ϵ-differential privacy. With this design, the principal can achieve any given accuracy objective by using the payment mechanism associated with an appropriate ϵ. In contrast to most of the existing work, which considers trusted principal and thus focuses on designing truthful mechanisms, this work is the first one to consider untrustworthy principal in private data collection and quality control mechanisms in such a scenario. We also show that the total expected payment of the designed mechanism at equilibrium is asymptotically optimal in the high data quality regime.

I. Introduction

With the rapid advancement of information technology, massive amounts of human-related data, such as health records, web browsing history, and opinions towards controversial issues, can be analyzed to uncover important findings in various areas, and to provide better understandings of human behaviors. However, the increasing concern of privacy poses a great challenge to this emerging field. Several serious privacy breach incidents [1]–[3] have made people more aware of the potential harm caused by privacy breach, and more cautious about giving away their data. Consequently, human-related data collection is impeded by the apparent obstacles of privacy concern and requires a systematic treatment.

We consider the problem of eliciting data from privacy-sensitive individuals. Specifically, an entity, who is named the *principal*, would like to collect data from individuals for big-data analytics. To incentivize individuals to participate and report informative data, the principal offers payments to participating individuals for their reported data. However, due to privacy concerns, an individual may be willing to report only a "noisy" version of the data for the sake of protecting her privacy (we will use "she" as a generic singular pronoun in this paper). The noise level is intimately tied with both the level of privacy protection and the quality of the reported data. We consider a model in which individuals take into account privacy loss during the data reporting stage, i.e., we do not assume the principal to be trustworthy, which differentiates our work from most of the existing work on privacy-aware surveys [4]–[10]. This model is meaningful in two aspects. First, after the witness of several privacy breach incidents, in which data holders like Netflix failed to protect the privacy of the individuals who contributed their data [2], individuals tend to consider the principal not trustworthy and prefer to control privacy by themselves. On the other hand, in some applications, such as collecting certain browsing history records to enhance the phishing and malware protection of web browsers [11], [12], the principal may not want to store individuals' original data either, to avoid subpoena risks. As a result, individuals are expected to report privacy-preserving data.

To quantify the privacy loss of individuals during data reporting, we consider the notion of (local) differential privacy [13]–[16]. Individuals need to randomize their data and then report the altered data for privacy protection measured by this notion. To make use of the reported data, the principal needs to know how the data has been randomized, i.e., the randomization strategies of individuals. However, this information is not verifiable, and individuals are not obligated to truthfully tell the principal what randomization strategies have been used. Then a challenge faced by the principal when collecting privacy-preserving data is: how to design the payment mechanism such that individuals randomize their data in a *predictable* way? The randomization procedure adds distortion to the data and thus degrades the quality of the reported data, whereas the accuracy objective for data analytics needs certain requirements on the data quality to be satisfied. Therefore, the above challenge can be taken a step further. How can the payment mechanism be designed such that individuals randomize their data with a *desired* level of quality? Successfully addressing this challenge will give the principal control over data quality when collecting privacy-preserving data from privacy-sensitive individuals.

In this paper, we take the first step towards addressing this

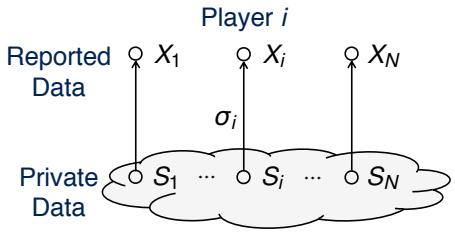

Fig. 1: Model of eliciting data from privacy-sensitive individuals. Each individual i has a private bit S_i and reports X_i, which is generated based on S_i using a randomized strategy.

important problem and consider the following model, which is illustrated in Fig. 1. There are N individuals and each individual i has a private bit S_i, e.g., her rating of a movie, which is either "good" or "bad" like in the rotten tomatoes website. The joint probability distribution of $S_1, S_2, \ldots, S_N$ is common knowledge. The principal is interested in learning the proportion of 1's in the private bits, which can be viewed as the popularity of a movie. This learning problem has been studied intensively in the literature (see, e.g., [4], [10], [17]). The principal uses a payment mechanism to determine the amount of payment to each individual based on their reported data $X_1, X_2, \ldots, X_N$. When an individual i uses an ϵ-differentially private randomization algorithm to generate her reported data X_i, the privacy loss incurred is ϵ, and her cost of privacy is a function of ϵ. The form of this function is also publicly known.

We study this problem with a game-theoretic approach, where we assume the individuals are strategic and hence the quality of data an individual reports is determined by her best response that takes into account both the payment and the privacy loss. A primary goal of the principal is to design a payment mechanism in which an individual's best response (or the Nash equilibrium of the game) has the desired level of quality. To design such a payment mechanism, we borrow ideas from the peer prediction method [18], which makes use of the correlation among private data (which is called signals in their context) to induce truthful reporting from individuals who have no privacy concern. We should caution that different from the peer prediction method, the privacy concern of individuals in this study fundamentally changes the structure of the game and gives the following distinctive features to our problem. First, since the notion of differential privacy is adopted, the privacy loss of an individual i is determined by both the strategy for $S_i = 1$ and that for $S_i = 0$. Therefore, when choosing the randomization strategy, an individual needs to perform joint optimization over the two possibilities and make a contingent plan. Second, the mechanism in this paper is not intended to elicit truthful data reporting. The principal is satisfied with the data quality as long as the accuracy objective can be achieved. In fact, truthful reporting may even not be preferred since it would otherwise cost the principal unnecessary additional payments. Consequently, when we build this study upon the peer prediction method, the prediction should be made on the randomized data instead of the original data.

Taking these features into consideration, we design a payment mechanism in which the randomized response strategy [19] that generates the reported data by flipping the private bit with probability $\frac{1}{e^\epsilon+1}$, where $\epsilon > 0$, proves to be an equilibrium. This equilibrium strategy is ϵ-differentially private, so the collected data itself is privacy preserving. By adjusting the corresponding parameter in the mechanism, the principal can control the privacy level ϵ and thus control the data quality to achieve any given accuracy objective. In contrast to most of the existing work, which considers a trusted principal and thus focuses on designing truthful mechanisms, our designed mechanism addresses individuals' privacy concern where the principal may not be trusted, and is the first one that considers quality control in such a scenario to suit the principle's accuracy objective.

II. Related Work

Most previous work on privacy-aware surveys [4]–[10], [20] assumes that there is a trusted entity that plays the role of the principal in this paper, who is expected to only use individuals' data in the announced way. The private data is either already kept by the principal, or is elicited using mechanisms that are designed with the aim of truthfulness.

In the seminal work by Ghosh and Roth [4], individuals' data is known to the principal, but when using the data for analysis, the analyst needs to pay the individuals to compensate their cost of privacy loss caused by the data usage, where each individual's privacy cost is modeled as a linear function of ϵ if her data is used in an ϵ-differentially private manner. The goal of the mechanism design there is to elicit truthful bids of individuals' cost functions, i.e., the coefficients. Subsequent work [5]–[7], [9] explores models for individuals' valuation of privacy, especially the correlation between the coefficients and the private bits.

This line of work has been extended to the scenario that the data is not available yet and need to be reported by the individuals to the trusted principal [8], [10]. Notably, Ghosh, Ligett and Roth [10] study the model in which the collected data is non-verifiable. The goal of the mechanism design there is to incentivize truthful data reporting from individuals. The problem of designing truthful mechanisms with players who have explicit value for privacy has also been studied in settings other than eliciting data for statistical analysis [21]–[23]. For example, Chen, Chong, Kash et al. [22] design an election mechanism that truthfully elicits votes from privacy-sensitive individuals and maximizes social welfare. For more work on the interplay between differential privacy and mechanism design, please see [24] for a comprehensive survey by Pai and Roth. Ours differs from these existing work by considering an untrustworthy principal.

The local model of differential privacy, which is a generalization of randomized response [19] and is formalized in [15], has been studied in the literature [13], [14], [16], [25]–[29]. In practice, Google's Chrome web browser has implemented the RAPPOR mechanism [11], [12] to collect users' data, which guarantees that only limited privacy of users will be leaked

by using randomized response in a novel manner. However, users may still not be willing to report data in the desired way due to the lack of an incentive mechanism.

III. Model

In this section, we present our model for the problem of eliciting data from privacy-sensitive individuals, which is illustrated Fig. 1. In this model, a principal collects data from individuals and offers payments to incentivize them to provide informative data. The payments can be monetary, or some privileges for individuals' accounts. For each individual, a cost for privacy loss, measured by the same unit as the payment, is incurred when reporting the data. Each individual strategically decides whether to report or not, and what to report, according to her own utility, which is the difference between the payment and the privacy cost. The individuals will be referred to as players in the remainder of this paper.

Consider a population of N players and denote the set of players by $\mathcal{N} = \{1, 2, \ldots, N\}$. As a standard notion in game theory, we denote all players except player i by "$-i$". Each player i has a private bit S_i, and let $S = (S_1, S_2, \cdots, S_N)$. The joint probability distribution of $S_1, S_2, \cdots, S_N$ is common knowledge. We assume that this distribution is symmetric over players; i.e., for any binary sequence $s \in \{0,1\}^N$ and any of its permutations s', $\mathbb{P}(S = s) = \mathbb{P}(S = s')$.

Let X_i denote the data reported by player i and let $X = (X_1, X_2, \ldots, X_N)$. The acceptable values for reported data are 0 and 1. So X_i takes values in the set $\mathcal{X} = \{0, 1, \perp\}$, where $\perp$ indicates that player i declines to participate. A strategy of player i for data reporting is a mapping $\sigma_i \colon \{0,1\} \to \mathcal{D}(\mathcal{X})$, where $\mathcal{D}(\mathcal{X})$ is the set of probability distributions on $\mathcal{X}$. Let $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_N)$. The strategy σ_i prescribes a distribution to X_i for each possible value of S_i, which defines the conditional distribution of X_i given S_i. Since we will discuss different strategies of player i, we let $\mathbb{P}_{\sigma_i}(X_i = x_i \mid S_i = s_i)$, where $x_i \in \mathcal{X}$ and $s_i \in \{0,1\}$, denote the conditional probabilities given strategy σ_i.

The principal is interested in learning the proportion of 1's in $S_1, S_2, \ldots, S_N$, i.e., $\bar{S} = \frac{1}{N}\sum_{i=1}^{N} S_i$. Let $\hat{\mu}$ be an estimate of $\bar{S}$ from the reported data $X_1, X_2, \ldots, X_N$. Then we measure the accuracy of $\hat{\mu}$ by the following definition, which has been used in the literature (e.g., [4], where a fixed number $\frac{1}{3}$ is used instead of δ).

Definition 1. An estimate $\hat{\mu}$ of $\bar{S}$ is (α, δ)-*accurate* if $|\bar{S} - \hat{\mu}| \le \alpha$ holds with probability at least $1 - \delta$.

The principal uses a payment mechanism $R \colon \mathcal{X}^N \to \mathbb{R}^N$ to determine the amount of payment to each individual, where $R_i(x)$ is the payment to player i when the reported data is $X = x$. We are interested in payment mechanisms in which the payment to each player is nonnegative, i.e., $R_i(x) \ge 0$ for any player i and any $x \in \mathcal{X}^N$, which we call *nonnegative mechanisms*. This constraint is motivated by the fact that in many practical applications such as surveys, the principal has no means to charge users and can only use payments to incentivize user participation.

To evaluate the cost of privacy loss during data reporting, a quantitative measure of privacy is needed. We define the privacy loss incurred when a strategy is used to be the level of (local) differential privacy, given in the following definition [13]–[16].

Definition 2. The *level of (local) differential privacy*, or simply the *privacy level*, of a strategy σ_i, denoted by $\zeta(\sigma_i)$, is defined as

$$\zeta(\sigma_i) = \max\left\{ \ln\left(\frac{\mathbb{P}_{\sigma_i}(X_i \in \mathcal{E} \mid S_i = s_i)}{\mathbb{P}_{\sigma_i}(X_i \in \mathcal{E} \mid S_i = 1 - s_i)} \right) : \mathcal{E} \subseteq \{0, 1, \perp\}, s_i \in \{0,1\} \right\}, \quad (1)$$

where we follow the convention that $0/0 = 1$.

We assume that different players experience the same cost of privacy loss if their strategies have the same privacy level. Thus, we model each player's cost of privacy loss by a function g of the privacy level. We call g the *cost function* and the cost the *privacy cost*. We say the cost function g is *proper* if it satisfied the following three conditions:

$$g(\xi) \ge 0, \quad \forall \xi \ge 0, \quad (2)$$

$$g(0) = 0, \quad (3)$$

$$g \text{ is non-decreasing}, \quad (4)$$

where (2) means that a privacy cost is nonnegative, (3) means that the privacy cost is 0 when the reported data is independent of the private data, and (4) means that the privacy cost will not decrease when the privacy loss becomes larger. In this paper, we will focus on a proper cost function that is convex and continuously differentiable. With a little abuse of notation, we also use $g(\sigma_i)$ to denote $g(\zeta(\sigma_i))$, which is the privacy cost to player i when the strategy σ_i is used.

The utility of each player is the difference between her payment and her privacy cost. We assume that the players are risk neutral, i.e., they are interested in maximizing their expected utility. We focus on the Nash equilibrium of a payment mechanism, where each player has no incentive to unilaterally change her strategy given other players' strategies. Formally, a Nash equilibrium in our model is defined as follows.

Definition 3. A strategy profile σ is a *Nash equilibrium* in a payment mechanism R if for any player i and any strategy σ_i',

$$\mathbb{E}_{\sigma}[R_i(X) - g(\sigma_i)] \ge \mathbb{E}_{(\sigma_i', \sigma_{-i})}[R_i(X) - g(\sigma_i')], \quad (5)$$

where the expectation is over the reported data X, and the subscripts σ and (σ_i', σ_{-i}) indicate that X is generated by the strategy profile σ and (σ_i', σ_{-i}), respectively.

IV. A Payment Mechanism for Quality Control

We wish to design mechanisms such that the quality of the collected data in equilibrium is controllable. Then the principal can achieve her accuracy objective by adjusting parameters in the mechanism. In this section, we present our

design of the payment mechanism. Consider the following payment mechanism $R^{(N,\epsilon)}$ for collecting privacy-preserving data from N players, parameterized by a data quality parameter ϵ, where $N \geq 2$ and $\epsilon > 0$.

The payment mechanism $R^{(N,\epsilon)}$:

1) Each player reports her data (which can also be the decision of not participating).
2) For non-participating players, the payment is zero.
3) If there is only one participant, pay zero to this participant. Otherwise, for each participating player i, arbitrarily choose another participating player j and pay player i according to X_i and X_j as follows:

$$R_i^{(N,\epsilon)}(X) = \frac{g'(\epsilon)(e^\epsilon+1)^2}{2e^\epsilon} A_{X_i,X_j}, \tag{6}$$

where parameters $A_{1,1}, A_{0,0}, A_{0,1}, A_{1,0}$ are calculated in the next section.

After the collection of data, the principal estimates $\bar{S} = \frac{1}{N}\sum_{i=1}^N S_i$ by

$$\hat{\mu} = \frac{e^\epsilon+1}{e^\epsilon-1}\left(\frac{1}{n}\sum_{i:\ X_i\neq\perp} X_i\right) - \frac{1}{e^\epsilon-1}, \tag{7}$$

where n is the number of participants.

A. Payment Parameterization

Recall that we assume that the joint distribution of $S_1, S_2, \ldots, S_N$ is symmetric over players. As a consequence, the private bits of the players have the same marginal distribution. Denote this marginal distribution as follows:

$$P_1 = \mathbb{P}(S_i = 1), \quad P_0 = \mathbb{P}(S_i = 0). \tag{8}$$

Due to symmetry, the marginal distribution of any two private bits S_i and S_j with $i \neq j$ does not depend on the specific identities i and j either. Denote the marginal distribution of S_i and S_j with $i \neq j$ as follows:

$$\begin{aligned} &P_{1,1} = \mathbb{P}(S_i = 1, S_j = 1), \quad P_{0,0} = \mathbb{P}(S_i = 0, S_j = 0), \\ &P_{0,1} = \mathbb{P}(S_i = 0, S_j = 1) = P_{1,0} = \mathbb{P}(S_i = 1, S_j = 0). \end{aligned} \tag{9}$$

We further define a constant D as follows:

$$\begin{aligned} D &= \mathbb{P}(S_j = 1, S_i = 1)\mathbb{P}(S_j = 0, S_i = 0) \\ &\quad - \mathbb{P}(S_j = 0, S_i = 1)\mathbb{P}(S_j = 1, S_i = 0) \\ &= P_{1,1}P_{0,0} - P_{0,1}P_{1,0}, \end{aligned} \tag{10}$$

which can be verified to equal to the covariance of S_i and S_j. We assume that $D \neq 0$, which is equivalent to the case that S_i and S_j are not independent for any two distinct players i and j (See Appendix A for the proof of the equivalence).

The parameters $A_{1,1}, A_{0,0}, A_{0,1}, A_{1,0}$ used in the payment mechanism $R^{(N,\epsilon)}$ are defined as follows:

- If $D > 0$,

$$A_{1,1} = \frac{(e^\epsilon+1)^2}{e^{2\epsilon}-1}\frac{1}{D}\left(\frac{1}{e^\epsilon+1}P_1 + \frac{e^\epsilon}{e^\epsilon+1}P_0\right), \tag{11}$$

$$A_{0,0} = \frac{(e^\epsilon+1)^2}{e^{2\epsilon}-1}\frac{1}{D}\left(\frac{e^\epsilon}{e^\epsilon+1}P_1 + \frac{1}{e^\epsilon+1}P_0\right), \tag{12}$$

$$A_{0,1} = 0, \tag{13}$$

$$A_{1,0} = 0. \tag{14}$$

- If $D < 0$,

$$A_{1,1} = 0, \tag{15}$$

$$A_{0,0} = 0, \tag{16}$$

$$A_{0,1} = -\frac{(e^\epsilon+1)^2}{e^{2\epsilon}-1}\frac{1}{D}\left(\frac{1}{e^\epsilon+1}P_1 + \frac{e^\epsilon}{e^\epsilon+1}P_0\right), \tag{17}$$

$$A_{1,0} = -\frac{(e^\epsilon+1)^2}{e^{2\epsilon}-1}\frac{1}{D}\left(\frac{e^\epsilon}{e^\epsilon+1}P_1 + \frac{1}{e^\epsilon+1}P_0\right). \tag{18}$$

From the above definition of these parameters we can see the intuition behind the design of mechanism $R^{(N,\epsilon)}$. When the private bits of two players are positively correlated ($D > 0$), they tend to be the same. Thus, the mechanism rewards agreement on the reported data to encourage informative data reporting. Similarly, when the private bits of two players are negatively correlated ($D < 0$), they tend to be different, and thus correspondingly, the mechanism rewards disagreement to encourage informative data reporting. However, the more informative the reported data is, the more privacy cost a player will experience. This tension will make each player choose a compromise, which is telling truth to some extent.

B. Nash Equilibrium

Theorem 1. *The strategy profile, consisting of the following strategy of player i that is denoted by σ_i^*, is a Nash equilibrium under the payment mechanism $R^{(N,\epsilon)}$:*

$$\begin{aligned} &\mathbb{P}_{\sigma_i^*}(X_i = 1 \mid S_i = 1) = \mathbb{P}_{\sigma_i^*}(X_i = 0 \mid S_i = 0) = \frac{e^\epsilon}{e^\epsilon+1}, \\ &\mathbb{P}_{\sigma_i^*}(X_i = 0 \mid S_i = 1) = \mathbb{P}_{\sigma_i^*}(X_i = 1 \mid S_i = 0) = \frac{1}{e^\epsilon+1}, \\ &\mathbb{P}_{\sigma_i^*}(X_i = \perp \mid S_i = 1) = \mathbb{P}_{\sigma_i^*}(X_i = \perp \mid S_i = 0) = 0, \end{aligned} \tag{19}$$

i.e., each player generates her reported data by flipping the private bit with probability $\frac{1}{e^\epsilon+1}$.

Proof. The proof is deferred to our technical report [30]. □

By Theorem 1, the parameter ϵ of the payment mechanism $R^{(N,\epsilon)}$ plays two roles in the equilibrium σ^*. On one hand, the strategy each player uses to randomize her data is ϵ-differentially private. Therefore, the parameter ϵ controls how much privacy each player is willing to trade for payment. On the other hand, the parameter ϵ describes the quality of the reported data of each player i, since ϵ controls the probability that the reported data is the same as the true private data as follows:

$$\mathbb{P}_{\sigma_i^*}(X_i = S_i) = \frac{e^\epsilon}{e^\epsilon+1}. \tag{20}$$

Therefore, the larger ϵ is, the more privacy each player is willing to sell, and the higher data quality the principal obtains. With the payment mechanism $R^{(N,\epsilon)}$, the principal is not only able to know how the data has been randomized, but also able to control the quality of the collected data.

C. Estimation Accuracy

In this section, we discuss how the principal should choose the parameter ϵ to achieve the accuracy objective of estimating $\bar{S}$.

Theorem 2. *For any α, δ with $\alpha > 0$ and $0 < \delta < 1$, if*

$$\epsilon \geq \ln\left(2 + \frac{1}{N\alpha^2\delta}\right), \tag{21}$$

then in the equilibrium σ^* of the payment mechanism $R^{(N,\epsilon)}$, the estimate $\hat{\mu}$ given in (7) *is (α, δ)-accurate.*

Proof. The proof is deferred to our technical report [30]. □

Since the parameter ϵ of the payment mechanism $R^{(N,\epsilon)}$ describes the quality of the collected data in the equilibrium σ^*, intuitively, the principal can achieve higher accuracy objective by increasing ϵ. Theorem 2 confirms this intuition. For an accuracy objective (α, δ), the smaller α and δ are, the higher accuracy is required to achieve according to the definition of accuracy in Definition 1. However, no matter how high the accuracy objective is, by Theorem 2, the principal can always achieve it by choosing large enough ϵ, i.e., good enough data quality.

D. Asymptotic Optimality in the High Quality Regime

From the principal's perspective, the strategy profile σ^* given in Theorem 1 is very attractive. When players follow σ^*, the quality of the collected data can be controlled by a single parameter ϵ, and $\bar{S}$ can be estimated by the simple estimator $\hat{\mu}$. In this section, we focus on nonnegative payment mechanisms in which σ^* forms a Nash equilibrium. We study the optimality of the proposed mechanism in terms of the total expected payment needed to collect data with a given quality level ϵ. We first derive an lower bound on the total expected payment of a nonnegative payment mechanism in which σ^* is an equilibrium. Then we compare the expected payment of the proposed mechanism with this lower bound and show that the proposed mechanism is asymptotically optimal in the high quality regime, i.e., as ϵ goes to infinity.

Proposition 1. *For any nonnegative payment mechanism R in which σ^* is a Nash equilibrium, the total expected payment at σ^* is lower bounded, given as follows*

$$\mathbb{E}_{\sigma^*}\left[\sum_{i=1}^{N} R_i(X)\right] \geq Ng'(\epsilon)(e^\epsilon + 1). \tag{22}$$

Proof. This lower bound is obtained through necessary conditions for the best response to have a noise level ϵ. The detailed proof is deferred to our technical report [30]. □

Therefore, to have an equilibrium at σ^*, a nonnegative payment mechanism needs to pay at least $Ng'(\epsilon)(e^\epsilon + 1)$ to the players. In the asymptotic regime that ϵ goes to infinity, this lower bound is on the order of $\mathcal{O}(g'(\epsilon)e^\epsilon)$.

In the equilibrium σ^* of the payment mechanism $R^{(N,\epsilon)}$, the total expected payment is given by

$$\begin{aligned}
&\mathbb{E}_{\sigma^*}\left[\sum_{i=1}^{N} R_i^{(N,\epsilon)}(X)\right] \\
&= Ng'(\epsilon)(e^\epsilon + 1) && (23)\\
&\quad + \frac{Ng'(\epsilon)(e^\epsilon+1)^2}{2e^\epsilon}\frac{(e^\epsilon+1)^2}{e^{2\epsilon}-1}\frac{1}{|D|} && (24)\\
&\quad \cdot \left(\frac{e^{2\epsilon}}{(e^\epsilon+1)^2}P_{0,1} + \frac{e^\epsilon}{(e^\epsilon+1)^2}(P_1^2 + P_0^2)\right. && (25)\\
&\quad \left. + \frac{1}{(e^\epsilon+1)^2}(P_1P_{1,1} + P_0P_{0,0})\right), && (26)
\end{aligned}$$

which can be obtained from the proof of Theorem 1.

In the asymptotic regime that ϵ goes to infinity, the total expected payment of mechanism $R^{(N,\epsilon)}$ is dominated by the first term, which is identical to the lower bound $Ng'(\epsilon)(e^\epsilon + 1)$, so the mechanism is asymptotically optimal in the high-quality regime.

V. Conclusion

In this paper we have shown how to design the payment mechanism to achieve quality control when collecting data from privacy-sensitive individuals. We considered a model in which individuals do not trust the principal and take into account a privacy cost that depends on the level of the (local) differential privacy of the data reporting strategy. Due to privacy concerns, an individual may be only willing to report a noisy version of the private data, which degrades the quality of the collected data. Our proposed mechanism incentives individuals to use a randomized response strategy with a desired noise level in the Nash equilibrium. This strategy generates the reported data by flipping the private data with probability $\frac{1}{e^\epsilon+1}$, where $\epsilon > 0$ is a parameter of the mechanism. Therefore, the quality of the collected data is controllable by adjusting ϵ. We also showed that the total expected payment of the designed mechanism is asymptotically optimal in the high quality regime. Note that the model of the private data in this work is a very general one. Considering some specific but well motivated structure for the model of the private data to find better mechanisms is an exciting direction for future work.

VI. Acknowledgement

This work was supported in part by the NSF under Grant ECCS-1255425.

Appendix A

In this section we prove that $D \neq 0$ is equivalent to the statement S_i and S_j are not independent for any two distinct players i and j. The direction that $D \neq 0$ implies dependence is obvious since D is the covariance of S_i and S_j.

For the other direction, suppose by contradiction that $D = 0$. Consider any two distinct players i and j. Recall that

$$P_1 = \mathbb{P}(S_i = 1), \quad P_0 = \mathbb{P}(S_i = 0).$$

First notice that $P_1 \neq 0$ and $P_0 \neq 0$ since otherwise S_i and S_j are independent. Then $D = 0$ implies that

$$\begin{aligned} &\mathbb{P}(S_j = 1 \mid S_i = 1)\mathbb{P}(S_j = 0 \mid S_i = 0) \\ &= \mathbb{P}(S_j = 0 \mid S_i = 1)\mathbb{P}(S_j = 1 \mid S_i = 0). \end{aligned} \tag{27}$$

Since $\mathbb{P}(S_j = 0 \mid S_i = 1) = 1 - \mathbb{P}(S_j = 1 \mid S_i = 1)$ and $\mathbb{P}(S_j = 1 \mid S_i = 0) = 1 - \mathbb{P}(S_j = 0 \mid S_i = 0)$, (27) further implies that

$$\begin{aligned} \mathbb{P}(S_j = 1 \mid S_i = 1) &= 1 - \mathbb{P}(S_j = 0 \mid S_i = 0) \\ &= \mathbb{P}(S_j = 1 \mid S_i = 0). \end{aligned}$$

Similarly,

$$\mathbb{P}(S_j = 0 \mid S_i = 1) = \mathbb{P}(S_j = 0 \mid S_i = 0).$$

Therefore, S_i and S_j are independent, which contradicts with the assumption that they are not independent. This completes the proof.

REFERENCES

[1] M. Barbaro and T. Zeller, "A face is exposed for AOL searcher no. 4417749," *New York Times*, Aug. 2006.

[2] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *Proc. IEEE Symp. Security and Privacy (SP)*, Oakland, CA, 2008, pp. 111–125.

[3] L. Sweeney, A. Abu, and J. Winn, "Identifying participants in the personal genome project by name (A re-identification experiment)," *arXiv:1304.7605 [cs.CY]*, Apr. 2013.

[4] A. Ghosh and A. Roth, "Selling privacy at auction," in *Proc. ACM Conf. Electronic Commerce (EC)*, San Jose, CA, 2011, pp. 199–208.

[5] L. K. Fleischer and Y. Lyu, "Approximately optimal auctions for selling privacy when costs are correlated with data," in *Proc. ACM Conf. Electronic Commerce (EC)*, Valencia, Spain, 2012, pp. 568–585.

[6] K. Ligett and A. Roth, "Take it or leave it: Running a survey when privacy comes at a cost," in *Proc. Int. Conf. Internet and Network Economics (WINE)*, Liverpool, UK, 2012, pp. 378–391.

[7] A. Roth and G. Schoenebeck, "Conducting truthful surveys, cheaply," in *Proc. ACM Conf. Electronic Commerce (EC)*, Valencia, Spain, 2012, pp. 826–843.

[8] A. Ghosh and K. Ligett, "Privacy and coordination: Computing on databases with endogenous participation," in *Proc. ACM Conf. Electronic Commerce (EC)*, 2013, pp. 543–560.

[9] K. Nissim, S. Vadhan, and D. Xiao, "Redrawing the boundaries on purchasing data from privacy-sensitive individuals," in *Proc. Conf. Innovations in Theoretical Computer Science (ITCS)*, Princeton, NJ, 2014, pp. 411–422.

[10] A. Ghosh, K. Ligett, A. Roth, and G. Schoenebeck, "Buying private data without verification," in *Proc. ACM Conf. Economics and Computation (EC)*, Palo Alto, CA, 2014, pp. 931–948.

[11] Ú. Erlingsson, V. Pihur, and A. Korolova, "RAPPOR: Randomized aggregatable privacy-preserving ordinal response," in *Proc. ACM SIGSAC Conf. Computer and Communication Security (CCS)*, Scottsdale, AZ, 2014, pp. 1054–1067.

[12] G. C. Fanti, V. Pihur, and Ú. Erlingsson, "Building a RAPPOR with the unknown: Privacy-preserving learning of associations and data dictionaries," *arXiv:1503.01214 [cs.CR]*, 2015.

[13] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. Conf. Theory of Cryptography (TCC)*, New York, NY, 2006, pp. 265–284.

[14] C. Dwork, "Differential privacy," in *Proc. Int. Conf. Automata, Languages and Programming (ICALP)*, Venice, Italy, 2006, pp. 1–12.

[15] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?" *SIAM J. Comput.*, vol. 40, no. 3, pp. 793–826, May 2011.

[16] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3–4, pp. 211–407, Aug. 2014.

[17] A. Blum, K. Ligett, and A. Roth, "A learning theory approach to non-interactive database privacy," in *Proc. Ann. ACM Symp. Theory of Computing (STOC)*, Victoria, Canada, 2008, pp. 609–618.

[18] N. Miller, P. Resnick, and R. Zeckhauser, "Eliciting informative feedback: The peer-prediction method," in *Computing with Social Trust*, ser. Human–Computer Interaction Series. Springer London, 2009, pp. 185–212.

[19] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *J. Amer. Stat. Assoc.*, vol. 60, no. 309, pp. 63–69, Mar. 1965.

[20] P. Dandekar, N. Fawaz, and S. Ioannidis, "Privacy auctions for recommender systems," *ACM Trans. Econ. Comput.*, vol. 2, no. 3, pp. 12:1–12:22, Jul. 2014.

[21] D. Xiao, "Is privacy compatible with truthfulness?" in *Proc. Conf. Innovations in Theoretical Computer Science (ITCS)*, Berkeley, CA, 2013, pp. 67–86.

[22] Y. Chen, S. Chong, I. A. Kash, T. Moran, and S. Vadhan, "Truthful mechanisms for agents that value privacy," in *Proc. ACM Conf. Electronic Commerce (EC)*, Philadelphia, PA, 2013, pp. 215–232.

[23] Y. Chen, O. Sheffet, and S. Vadhan, "Privacy games," in *Int. Conf. Web and Internet Economics (WINE)*, vol. 8877, 2014, pp. 371–385.

[24] M. M. Pai and A. Roth, "Privacy and mechanism design," *SIGecom Exch.*, vol. 12, no. 1, pp. 8–29, Jun. 2013.

[25] J. Hsu, S. Khanna, and A. Roth, "Distributed private heavy hitters," in *Proc. Int. Conf. Automata, Languages and Programming (ICALP)*, Warwick, UK, 2012, pp. 461–472.

[26] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy, data processing inequalities, and statistical minimax rates," *arXiv:1302.3203v4 [math.ST]*, Feb. 2013.

[27] W. Wang, L. Ying, and J. Zhang, "On the relation between identifiability, differential privacy, and mutual-information privacy," in *Proc. Ann. Allerton Conf. Commununication, Control and Computing (Allerton)*, Monticello, IL, Sep. 2014, pp. 1086–1092.

[28] ——, "A minimax distortion view of differentially private query release," in *Asilomar Conf. Signals, Systems, and Computers*, 2015, to appear.

[29] R. Bassily and A. Smith, "Local, private, efficient protocols for succinct histograms," in *Proc. Ann. ACM Symp. Theory of Computing (STOC)*, Portland, OR, 2015, pp. 127–135.

[30] W. Wang, L. Ying, and J. Zhang, "A game-theoretic approach to quality control for collecting privacy-preserving data," Arizona State University, Tech. Rep., Jul. 2015.