Buying Data from Privacy-Aware Individuals: The Effect of Negative Payments

Weina Wang, Lei Ying, and Junshan Zhang

School of Electrical, Computer and Energy Engineering, Arizona State University
Tempe, AZ 85281, USA
{weina.wang, lei.ying.2, junshan.zhang}@asu.edu

Abstract. We study a market model where a data analyst uses monetary incentives to procure private data from strategic data subjects/ individuals. To characterize individuals' privacy concerns, we consider a local model of differential privacy, where the individuals do not trust the analyst so privacy costs are incurred when the data is reported to the data analyst. We investigate a basic model where the private data are bits that represent the individuals' knowledge about an underlying state, and the analyst pays each individual according to all the reported data. The data analyst's goal is to design a payment mechanism such that at an equilibrium, she can learn the state with an accuracy goal met and the corresponding total expected payment minimized. What makes the mechanism design more challenging is that not only the data but also the privacy costs are unknown to the data analyst, where the costs reflect individuals' valuations of privacy and are modeled by "cost coefficients." To cope with the uncertainty in the cost coefficients and drive down the data analyst's cost, we utilize possible negative payments (which penalize individuals with "unacceptably" high valuations of privacy) and explore interim individual rationality. We design a family of payment mechanisms, each of which has a Bayesian Nash equilibrium where the individuals exhibit a threshold behavior: the individuals with cost coefficients above a threshold choose not to participate, and the individuals with cost coefficients below the threshold participate and report data with quality guarantee. By choosing appropriate parameters, we obtain a sequence of mechanisms, as the number of individuals grows large. Each mechanism in this sequence fulfills the accuracy goal at a Bayesian Nash equilibrium. The total expected payment at the equilibrium goes to zero; i.e., this sequence of mechanisms is asymptotically optimal.

1 Introduction

Exploiting human-related data such as medical records and financial data has created great value to the society. However, the ever-improving capability of data analysis in the advancing big data technology makes it possible to extract personal information undesirably, giving rise to technical barriers for data collection. In short, big data analytics is a double edged sword. This in turn necessitates incentivizing data subjects/individuals for providing quality data while preserving privacy.

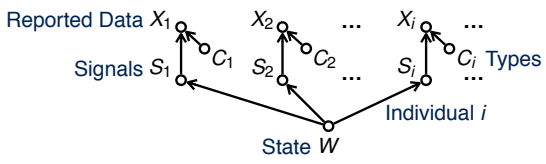

Fig. 1. System model. The data analyst is interested in the state W, which is a binary random variable. Each individual i has a private binary signal S_i and a type C_i that characterizes her valuation of privacy. $S_1, S_2, \ldots, S_N$ are conditionally i.i.d. given W. Individual i reports randomized data X_i, which is generated based on S_i and C_i.

In this paper, we consider a market model where a data analyst uses monetary incentives to procure private data from strategic data subjects/individuals. Specifically, the data analyst elicits data from a population of N individuals. Each individual i's private data is a binary *signal* S_i that reflects her knowledge about an underlying *state*, which is represented by a binary random variable W. Conditional on the state W, the signals are independently generated such that the probability for S_i to be the same as W is θ, where $0.5 < \theta < 1$. The data analyst is interested in learning W. This structure is illustrated in Figure 1.

To provide monetary incentives, the data analyst announces a mechanism, which is a function that determines the amounts of payments to individuals according to their reported data. Since an individual's payment may depend on others' reports, this payment mechanism induces a game among the individuals. Due to privacy concerns, an individual experiences a cost when she releases data to the analyst. Her payoff is the difference between the payment and the privacy cost. The goal of the data analyst in this setting is to design a mechanism to achieve a desired learning accuracy at an equilibrium in a cost-effective manner.

Privacy cost model. To quantify the privacy costs, we consider a local model of differential privacy (an introduction of which can be found in [10]). In this local model, the individuals do not trust the data analyst with their data, so they have to evaluate their privacy costs carefully when reporting data to the analyst. To control the privacy cost, we assume in the paper that an individual adds random noise to her data and reports the resulting perturbed version. Intuitively, the more "noisy" the reported data is, the more privacy is retained, and thus the less privacy cost is incurred. An individual will weigh the privacy cost against the payment to choose the best way of perturbing her data. In contrast, in a centralized model of data privacy with a trustworthy data analyst (e.g., [15]), the action of providing data to the analyst itself, whether truthful or not, does not cause privacy loss. There privacy costs are incurred when the analyst releases the outcome of the mechanism, so the individuals cannot control their privacy costs except choosing to participate or not.

We remark that the different privacy cost models make the structures of the mechanism design problem fundamentally different. In a centralized model, the design goal is to have a mechanism that elicits truthful data reporting and its outcome satisfies the promised privacy guarantee. However, in the local model

considered in this paper, truthfulness is no longer a focal design goal since it incurs high privacy costs to individuals that need to be compensated by payments. Instead, the data analyst seeks for mechanisms that cost-effectively elicit data with small enough perturbation, and consequently the analyst needs to manage equilibria consisting of noisy data reporting. Another major difference is that it is unnecessary to make the outcome of the mechanism guarantee privacy in the local model since the control of privacy remains in the hand of the individuals.

Unknown privacy valuations and the impact of negative payments. We consider the natural setting where different individuals may have different valuations of privacy and their valuations are unknown to the data analyst. In this model, the analyst is not able to tune the payments to the privacy costs, which may result in overpayments when some individuals' costs are lower than expected. This uncertainty can also introduce unwanted noise in the reported data. To see this, consider a mechanism that always pays a nonnegative amount of payment to a participant. For an individual whose valuation of privacy is very high, participating and reporting only noise is a better strategy than opting out since she may still receive some nonnegative amount of payment without incurring any privacy cost. Then the payment does not buy the analyst any useful information from this individual, and moreover, the analyst has to work with these unusable reports during data analysis.

With these observations, we consider payment mechanisms that are interim individually rational; i.e., the expected payoff of each individual in an equilibrium is nonnegative, although the realizations of the payments can be negative. In practice, this can model the scenario where there are repeated data collection (e.g., to learn the ratings of different movies), and in some rounds the payments received by the individual may be negative, but in the long run, the average payoff will be nonnegative. Negative payments can be utilized to "filter out" individuals with high privacy costs; i.e., we design the mechanism such that their expected payoff is negative if they report only noise. This saves the data analyst's payments on poor quality data and simplifies the data analysis. We will see that we can actually drive the total cost to zero for the data analyst as the population size becomes large.

We remark that one possible approach to implement negative payments is to let the data analyst set up an online payment system using virtual currency or credits. Instead of paying real money to an individual every time she reports a data, virtual currency or credits can be added to or reduced from the user's account. An individual can be paid weekly or monthly with real dollars. Since the expected payment is nonnegative, the real-dollar payment over a long time period is nonnegative with a high probability. We remark that negative payments may not be feasible in many scenarios. The focus of this paper is to reveal the fundamental benefit of negative payments to the data analyst when feasible.

Main Results. With the above formulation, we are interested in the following intriguing questions: (1) How will the individuals behave to reconcile the conflict between privacy and rewards? (2) How should the data analyst design the mechanism such that she can achieve her learning goal cost-effectively?

Let $X_1, X_2, \ldots, X_N$ denote the reported bits of the individuals. We model the privacy cost of an individual as a function of her privacy loss, which is measured by the level of (local) differential privacy [9,8] of her data reporting strategy. This cost function of individual i is characterized by her *type* C_i: when individual i reports data with a (local) differential privacy level of ϵ after observing her type $C_i = c_i$, her privacy loss is ϵ and the corresponding privacy cost is $c_i\epsilon$. The type of an individual is also called her cost coefficient due to this linear form. We assume that the types are i.i.d. and an individual's type is independent from her private data, which is applicable to the scenario where an individual's valuation of privacy is intrinsic and thus is not affected by the specific private data she has. The reported data and cost coefficients are also illustrated in Figure 1. We remark that both the settings where an individual's valuation of privacy is independent (e.g., [14]) and correlated (e.g., [15]) with her private data have been studied in the literature. We further assume that it is possible for individuals to have valuations arbitrarily close to zero. In this paper, the prior distribution of the state, signals and types is public information. However, neither the private signals nor the types are known to the data analyst.

Our main result is centered around constructing a family of payment mechanisms indexed by parameters, which provide answers to the proposed questions from the following perspectives.

– **Behavior of individuals with privacy concerns.** We show that the individuals exhibit a threshold behavior in a Bayesian Nash equilibrium of the proposed mechanism: the individuals with cost coefficients above a threshold c_{th} opt out, and the individuals with cost coefficients below c_{th} participate and report data with a privacy level no smaller than ϵ, where c_{th} and ϵ are parameters of the mechanism. Since a larger privacy level means that the data is less perturbed, the data analyst can incentivize the participants to limit the perturbation to a desired extent by choosing an appropriate ϵ. It can be seen from this result that this mechanism resolves the otherwise nuisance that individuals with high privacy costs may participate and report only noise: they are "filtered out", and the "remaining" participants all report data with quality guarantee.

– **Tradeoff between learning accuracy and cost.** We show that as the population size grows to infinity, the data analyst can learn the underlying state with arbitrarily small overall probability of error, with the total expected payment at the Bayesian Nash equilibrium going to zero. That is to say, if the data analyst can recruit a large number of individuals, she can choose appropriate parameters to fulfill her learning goal and in the meanwhile drive her cost to zero at a Bayesian Nash equilibrium. Since the total equilibrium expected payment of any mechanism is nonnegative due to individual rationality, this implies that the designed mechanism with properly chosen parameters asymptotically minimizes the cost for achieving any accuracy goal.

Related Work. Market approaches for collecting data from privacy-aware individuals have led to a fruitful line of work [16,13,20,26,14,35,3,24,15,5,29,32]. These papers except [5,29,32] adopt the centralized model for privacy. The seminal work by Ghosh and Roth [16] and a line of subsequent work [13,20,26,14,24]

considered the setting where the private data is verifiable so the individuals cannot misreport data, but they can strategically report their privacy costs. A recent work [5] considered a model where a data analyst procures possibly noisy estimates (data) from data providers. This can be thought of as a local privacy model, but still the data is verifiable. The setting of the work [35,3,15,29,32] is more similar to ours, where the individual have the option of misreporting data. The work [35,3,15] considered the centralized privacy model, where revealing data to the data analyst does not incur privacy costs. Then strategically reporting data can alter the individuals' payments but does affect their privacy costs. The work [29,32] considered the local privacy model but assumed the privacy cost functions are known to the data analyst. Our work studies this problem in a local privacy model, where neither the data nor the privacy cost functions are known. The mechanism thus needs to deal with the uncertainty in both sources and work with noisy reports.

The broader field of the interplay between differential privacy and mechanism design, first studied by McSherry and Talwar [22], is surveyed in [25]. The behavior of individuals with privacy concerns has been studied in [4], which investigates the types of games in which strategic individuals truthfully follow randomized response. The market approach for collecting private data also shares some structural similarity with the problem of information elicitation (e.g., [23]), especially the effort elicitation in the context of crowdsourcing (e.g., [6,34,2,21]), where effort, instead of privacy concerns, affects the quality of the data and the cost of the individuals.

The local model of differential privacy, which generalizes the randomized response [33], has been studied in the literature [9,8,19,17,7,10,4,18,28,30,1,27]. The hypothesis testing formulation in our paper is similar to a setting in [18], where the authors find an optimal locally differentially private privatization mechanism that maximizes the statistical discrimination of the hypotheses. In practice, Google's Chrome web browser has implemented the RAPPOR mechanism [11,12] to collect users' data using a locally differentially private protocol.

2 Model

We study the setting in which the data analyst is interested in learning an underlying state W, represented by a binary random variable. Consider a set $[N] = \{1, 2, \ldots, N\}$ of individuals. Each individual i possesses a binary signal S_i, which is her private data, and reports data X_i, which takes values in $\mathcal{X} = \{0, 1, \perp\}$, with $\perp$ meaning "to opt out." The data analyst announces a payment mechanism $\boldsymbol{R}\colon \mathcal{X}^N \to \mathbb{R}^N$, which takes the reported data $\boldsymbol{X} = (X_1, \ldots, X_N)$ as input and produces $\boldsymbol{R}(\boldsymbol{X})$, where $R_i(\boldsymbol{X})$ is the payment to individual i. The model is illustrated in Figure 1. The payment mechanism induces a game among the individuals. The elements of the game are as follows.

Players. The players in this game are the individuals, who are self-interested, rational and risk-neutral. Following conventional game theory notation, we let "$-i$" denote all the individuals other than some given individual i.

Prior distributions. The state W follows a probability distribution given by the PMF P_W. We assume that $P_W(1) > 0$ and $P_W(0) > 0$. The individuals' signals $\boldsymbol{S} = (S_1, S_2, \ldots, S_N)$ reflect their knowledge about the state W. Conditional on the state W, the signals $S_1, S_2, \ldots, S_N$ are independently generated according to $\mathbb{P}(S_i = w \mid W = w) = \theta$ for $w \in \{0, 1\}$, where the parameter θ with $0.5 < \theta < 1$ is called the quality of signals. We refer to these conditional distributions as the signal structure of the model.

Types and strategies. An individual i's type C_i, also called her cost coefficient, characterizes her valuation of privacy. We will elaborate on the assumptions on the types when we introduce the payoff functions below. Roughly, an individual with larger C_i experiences more privacy cost for the same privacy loss. A data reporting strategy for individual i is a plan on what to report according to her signal S_i and her type C_i. Thus it is a mapping $\sigma_i \colon \{0, 1\} \times (0, +\infty) \to \mathcal{D}(\mathcal{X})$, where $\mathcal{D}(\mathcal{X})$ is the set of probability distributions on $\mathcal{X} = \{0, 1, \perp\}$, prescribing a distribution to the reported data X_i for each possible value pair of S_i and C_i. Therefore, the strategy corresponds to the set of conditional distributions of X_i given S_i and C_i. Since we will discuss different strategies of individual i, we denote these conditional probabilities by $\mathbb{P}_{\sigma_i}(X_i = x_i \mid S_i = s_i, C_i = c_i)$ for $x_i \in \{0, 1, \perp\}, s_i \in \{0, 1\}$, and $c_i \in (0, +\infty)$. Let $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)$, which is called a strategy profile. A strategy profile is said to be homogeneous if all the strategies in the profile are the same.

Payoff functions. The payoff of each individual is the difference between the payment she receives and her privacy cost. An individual experiences a cost due to the privacy loss during data reporting. Recall that we model the privacy cost of an individual as consisting of two components: privacy loss and a privacy cost function, where the privacy loss depends on her data reporting strategy and the privacy cost function represents her valuation of privacy. For an individual i, conditional on her type $C_i = c_i$, we measure individual i's privacy loss for reporting data with strategy σ_i by the privacy level defined as follows:

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

where we follow the convention that $0/0 = 1$. This measure of privacy loss is in the same vein as the local model of differential privacy [19,10], which views each individual's data as a database of size 1 and quantifies the privacy guarantee of her local randomizer by the differential privacy level. The difference here is that the strategy σ_i has another input C_i, since an individual can choose the way of perturbing her data according to her cost coefficient. Our measure of privacy loss is the differential privacy level of the strategy σ_i when C_i is given.

Then we model individual i's cost incurred by this privacy loss as a linear function with C_i as the coefficient, i.e., the cost can be written as $g(C_i, \sigma_i) = C_i \cdot \zeta(C_i, \sigma_i)$. We call g the privacy cost function.

We assume that the coefficients $C_1, C_2, \ldots, C_N$ are i.i.d. positive random variables with CDF F_C, and they are independent of W and $\boldsymbol{S}$. The randomness of these coefficients captures the data analyst's uncertainty of individuals' valuations of privacy. The independence assumption is applicable to the scenario

where individuals' valuations of privacy are intrinsic and thus are not affected by the specific private data they have. For ease of exposition, we further assume that F_C is a continuous function and $F_C(c) > 0$ for any $c > 0$, which means that it is possible for individuals to have an arbitrarily low valuation of privacy. Similar analysis can be carried out for other models for the types (but the resulting accuracy–payment relation may be different).

Mechanism Design. The data analyst cannot force an individual to report data with a specific strategy. However, the data analyst can design the payment mechanism to impact individuals' strategies to drive the individuals to act in a desired way since the individuals are rational, i.e., they will choose the strategies that benefit them most. We consider the Bayesian Nash equilibria in a payment mechanism, viewing C_i as individual i's type.

Definition 1. *A strategy profile $\boldsymbol{\sigma}$ is a* Bayesian Nash equilibrium *of a payment mechanism $\boldsymbol{R}$ if for any individual i, any $c_i > 0$ and any strategy σ_i',*

$$\mathbb{E}_{\boldsymbol{\sigma}}[R_i(\boldsymbol{X}) - g(C_i, \sigma_i) \mid C_i = c_i] \geq \mathbb{E}_{(\sigma_i', \sigma_{-i})}[R_i(\boldsymbol{X}) - g(C_i, \sigma_i') \mid C_i = c_i],$$

where the subscript $\boldsymbol{\sigma}$ and (σ_i', σ_{-i}) indicate that the distribution of $\boldsymbol{X}$ is determined by the strategy profile $\boldsymbol{\sigma}$ and (σ_i', σ_{-i}), respectively.

The data analyst is interested in learning the state W from the reported data $\boldsymbol{X}$, so she performs hypothesis testing between the two hypotheses $H_0 \colon W = 0$ and $H_1 \colon W = 1$. The learning accuracy is measured by the overall probability of error, denoted by p_e, which is $P_W(0) \cdot (\text{Type I error}) + P_W(1) \cdot (\text{Type II error})$. An accuracy goal can be written as $p_e \leq p_e^{\max}$ for some $p_e^{\max}$.

Then the data analyst aims to design a payment mechanism such that her accuracy goal can be fulfilled at a Bayesian Nash equilibrium and the corresponding total expected payment is minimized. It is easy to see that the equilibrium total expected payment is nonnegative in any mechanism due to the nonnegativity of privacy cost functions and individual rationality. In this mechanism design problem, the joint distribution $\mathcal{P}$ of the state W, the signal $\boldsymbol{S}$ and the cost coefficients, which can be represented by (P_W, θ, F_C), is common knowledge. The data analyst announces the form of the payment mechanism and then the individuals report data simultaneously. The reported data $\boldsymbol{X}$ is public. Each individual i's signal and type, S_i and C_i, are not observable to other individuals or the data analyst. No one has access to the state W.

3 Asymptotically Optimal Mechanisms

Theorem 1. *To achieve any accuracy goal of the data analyst, the total expected payment needed at an equilibrium is $o(1)$. Specifically, there exists a sequence of mechanisms, each of which is designed for a different population size N, such that the accuracy goal can be fulfilled at a Bayesian Nash equilibrium of every mechanism in the sequence, and the total expected payment goes to zero as the population size N goes to infinity; i.e., this sequence of mechanisms is asymptotically optimal.*

In the remainder of this section, we present the design of a family of payment mechanisms, parameterized by the population size N, the prior $\mathcal{P}$, a cost coefficient threshold parameter c_{th} and a data quality parameter ϵ. The asymptotically optimal sequence of mechanisms in Theorem 1 is given by a sequence of mechanisms within this family with properly chosen parameters. In particular, c_{th} is a threshold on cost coefficients such that an individual is expected to participate if her coefficient does not exceed the threshold; and ϵ is the target quality which is the level noise expected in the reported data. The formula for calculating c_{th} and ϵ will be presented in Section 5. Theorem 1 is a high level description of Theorem 3, which will be derived in the remainder of this paper.

Payment Mechanism $\boldsymbol{R}^{(N,\mathcal{P},c_{\text{th}},\epsilon)}$

1. Each individual reports her data (which can also be "to opt out").
2. Compute the number of participants n.
3. For non-participating individuals, the payment is zero.
4. If there is only one participant, the data analyst pays zero to this participant. Otherwise, for each participating individual i, compute the majority of other participants' reported data, denoted by M_{-i}. Then the data analyst pays individual i according to X_i and M_{-i} as follows:

$$R_i^{(N,\mathcal{P},c_{\text{th}},\epsilon)}(\boldsymbol{X}) = A_{X_i,M_{-i}}\frac{c_{\text{th}}(e^\epsilon+1)^2}{2e^\epsilon} + B_{M_{-i}}\left(\frac{c_{\text{th}}(e^\epsilon+1)}{e^\epsilon} + c_{\text{th}}\epsilon\right),$$

where $A_{1,1}, A_{0,1}, A_{1,0}, A_{0,0}, B_1, B_0$ are given below.

Next we define the coefficients $A_{1,1}$, $A_{0,1}$, $A_{1,0}$, $A_{0,0}$, B_1, B_0 used in the mechanism $\boldsymbol{R}^{(N,\mathcal{P},c_{\text{th}},\epsilon)}$ through a series of calculations. In a nutshell, $A_{1,1}$ and $A_{0,0}$ determine the reward part of the payment to an individual when her reported data matches the majority of others; similarly, $A_{0,1}$ and $A_{1,0}$ determine the penalty part of the payment to an individual when her reported data does not match the majority of others. They incentivize the individuals to report data that reveals certain amount of information about their private signals. The coefficients B_1 and B_0 offset the payments for the cases that the majority of others' reports is 1 and 0, respectively, to discourage the individuals with cost coefficients above threshold parameter c_{th} from participating. We remark that when an individual's reported data does not match with the majority of others, these coefficients make sure that the payment to this individual is negative.

The definition of the coefficients $A_{1,1}$, $A_{0,1}$, $A_{1,0}$, $A_{0,0}$, B_1, B_0 involves some intermediate quantities, the physical meanings of which will be given after we characterize a Bayesian Nash equilibrium of the mechanism in Section 4. Given a $c_{\text{th}} \in (0,+\infty)$ and $\epsilon \in (0,+\infty)$, for each $c_i \in (0, c_{\text{th}})$, we consider the following equation with variable ξ: $c_{\text{th}}(e^\epsilon+1)^2e^\xi = c_i e^\epsilon(e^\xi+1)^2$. It can be proved that this equation has a unique solution in $(0,+\infty)$. Let this solution define a function $\xi(c_i)$. Specifically,

$$\xi(c_i) = \ln\left(\frac{1}{\frac{1}{2} - \sqrt{\frac{1}{4} - \frac{c_i}{c_{\text{th}}}\frac{e^\epsilon}{(e^\epsilon+1)^2}}} - 1\right). \tag{1}$$

Let

$$\mu = \int_0^{c_{\mathrm{th}}} \frac{e^{\xi(c_i)}}{e^{\xi(c_i)}+1}\, dF_{C|C_i \le c_{\mathrm{th}}}(c_i), \quad \alpha = \theta\mu + (1-\theta)(1-\mu), \tag{2}$$

where $F_{C|C_i \le c_{\mathrm{th}}}$ is the conditional distribution of C_i given $C_i \le c_{\mathrm{th}}$.

Given that the number of participants is n with $n \ge 2$, we define the following quantities. Consider a random variable that follows the binomial distribution with parameters $n-1$ and α. Let $\beta^{(n)}$ denote the probability that this random variable is greater than or equal to $\lfloor \frac{n-1}{2} \rfloor + 1$. For convenience, we define the following quantity to deal with technical details:

$$\gamma^{(n)} = \begin{cases} 1 - \binom{n-1}{\frac{n-1}{2}} \alpha^{\frac{n-1}{2}} (1-\alpha)^{\frac{n-1}{2}} & \text{if } n-1 \text{ is even,} \\ 1 & \text{if } n-1 \text{ is odd.} \end{cases}$$

Let $P_{\ge 1} = 1 - (1 - F_C(c_{\mathrm{th}}))^{N-1}$, where F_C is the CDF of C_i. We define

$$\begin{aligned}
A_{1,1} &= \frac{P_W(1)\theta(1-\beta^{(n)}) + P_W(0)(1-\theta)(1-(\gamma^{(n)}-\beta^{(n)}))}{P_{\ge 1}P_W(1)P_W(0)(2\theta-1)(2\beta^{(n)}-\gamma^{(n)})}, \\
A_{0,1} &= -\frac{P_W(1)(1-\theta)(1-\beta^{(n)}) + P_W(0)\theta(1-(\gamma^{(n)}-\beta^{(n)}))}{P_{\ge 1}P_W(1)P_W(0)(2\theta-1)(2\beta^{(n)}-\gamma^{(n)})}, \\
A_{1,0} &= -\frac{P_W(1)\theta\beta^{(n)} + P_W(0)(1-\theta)(\gamma^{(n)}-\beta^{(n)})}{P_{\ge 1}P_W(1)P_W(0)(2\theta-1)(2\beta^{(n)}-\gamma^{(n)})}, \\
A_{0,0} &= \frac{P_W(1)(1-\theta)\beta^{(n)} + P_W(0)\theta(\gamma^{(n)}-\beta^{(n)})}{P_{\ge 1}P_W(1)P_W(0)(2\theta-1)(2\beta^{(n)}-\gamma^{(n)})}, \\
B_1 &= -\frac{P_W(1)(1-\beta^{(n)}) - P_W(0)(1-(\gamma^{(n)}-\beta^{(n)}))}{2P_{\ge 1}P_W(1)P_W(0)(2\beta^{(n)}-\gamma^{(n)})}, \\
B_0 &= \frac{P_W(1)\beta^{(n)} - P_W(0)(\gamma^{(n)}-\beta^{(n)})}{2P_{\ge 1}P_W(1)P_W(0)(2\beta^{(n)}-\gamma^{(n)})}.
\end{aligned}$$

4 Bayesian Nash Equilibrium

In this section, we first characterize the individuals' behavior at a Bayesian Nash equilibrium of the designed mechanism. The equilibrium behavior affects the quality of the reported data and the payments. Then we leverage the properties of the Bayesian Nash equilibrium to explain the physical meanings of the quantities defined during the construction of the mechanism in Section 3.

Theorem 2. *The mechanism $\boldsymbol{R}^{(N,\mathcal{P},c_{\mathrm{th}},\epsilon)}$ yields a Bayesian Nash equilibrium $\boldsymbol{\sigma}$, in which each individual i's strategy σ_i is described as follows:*

- *If $c_i > c_{\mathrm{th}}$, $\mathbb{P}_{\sigma_i}(X_i = \perp \mid S_i = s_i, C_i = c_i) = 1$ for any $s_i \in \{0,1\}$; i.e., if individual i's cost coefficient is larger than the parameter c_{th}, individual i declines to participate regardless of her signal.*

– *If* $c_i \le c_{\text{th}}$,

$$\mathbb{P}_{\sigma_i}(X_i = 1 \mid S_i = 1, C_i = c_i) = \mathbb{P}_{\sigma_i}(X_i = 0 \mid S_i = 0, C_i = c_i) = \frac{e^{\xi(c_i)}}{e^{\xi(c_i)} + 1},$$
$$\mathbb{P}_{\sigma_i}(X_i = 0 \mid S_i = 1, C_i = c_i) = \mathbb{P}_{\sigma_i}(X_i = 1 \mid S_i = 0, C_i = c_i) = \frac{1}{e^{\xi(c_i)} + 1},$$

where $\xi(c_i)$ *is defined in* (1)*; i.e., if individual* i*'s cost coefficient is no larger than the parameter* c_{th}*, individual* i *flips her signal with a probability depending on her cost coefficient to generate her reported data.*

The following corollary describes the quality of the reported data and each participant's expected payment at the Bayesian Nash equilibrium in Theorem 2.

Corollary 1. *For the mechanism* $\boldsymbol{R}^{(N,\mathcal{P},c_{\text{th}},\epsilon)}$*, consider the Bayesian Nash equilibrium* $\boldsymbol{\sigma}$ *given in Theorem 2.*

– *For each participating individual* i*,*

$$\mathbb{P}_{\sigma_i}(X_i = 1 \mid S_i = 1, \textit{ i participates}) = \mathbb{P}_{\sigma_i}(X_i = 0 \mid S_i = 0, \textit{ i participates}) = \mu,$$

where μ *is defined in* (2) *and* $\mu \ge \frac{e^{\epsilon}}{e^{\epsilon}+1}$*.*

– *The expected payment to each participating individual* i *is bounded as*

$$\mathbb{E}_{\boldsymbol{\sigma}}[R_i^{(N,\mathcal{P},c_{\text{th}},\epsilon)}(\boldsymbol{X}) \mid \textit{ i participates}] \le c_{\text{th}}(1 + e^{-\epsilon} + \epsilon).$$

The proofs of Theorem 2 and Corollary 1 are presented in the full version [31]. Theorem 2 and Corollary 1 show how individuals with high privacy costs are "filtered out" in the equilibrium by negative payments. In other words, they will decide not to participate because the expected payment is negative, which is a result of the possible negative payments in the proposed mechanism. The "remaining" individuals, i.e., participants, all report data with quality guarantee. The roles of the parameters c_{th} and ϵ in the designed mechanism $\boldsymbol{R}^{(N,\mathcal{P},c_{\text{th}},\epsilon)}$ are as follows: The parameter c_{th} works as a threshold on the cost coefficients for participation; The parameter ϵ gives a guarantee on the probability that a participant's reported data is the same as the signal, which measures the quality of the reported data. We remark that in this equilibrium, each individual's exact cost coefficient is not revealed to other.

The physical meanings of the quantities $\xi(c_i)$, μ, α, $\beta^{(n)}$, $\gamma^{(n)}$ and $P_{\ge 1}$ defined during the construction of the mechanism in Section 3 can be well explained at the Bayesian Nash equilibrium given in Theorem 2. The quantity $\xi(c_i)$ shows up in Theorem 2, characterizing the strategy σ_i of individual i when $c_i \le c_{\text{th}}$. It is the differential privacy level of σ_i given $C_i = c_i$ when $c_i \le c_{\text{th}}$. Now let us condition on the event that individual i participates, which, by Theorem 2, is equivalent to the event $C_i \le c_{\text{th}}$. The quantity μ shows up in Corollary 1, and it is the probability that individual i truthfully reports her signal, given whatever

the signal is. Then the quantity α is the probability that the reported data X_i is consistent with the state W, given whatever the state is. Conditional on the event that there are $n-1$ participants among the individuals other than individual i, where $n \geq 2$, the quantities β_n and $1-(\gamma_n-\beta_n)$ are the probabilities that the majority of these participants' reported data agrees with the state, given that the state is 1 and 0, respectively. Finally, the quantity $P_{\geq 1}$ is the probability that at least one individual among the individuals other than individual i participates.

5 Accuracy and Payment

In this section, we show that the data analyst can achieve any accuracy goal in the Bayesian Nash equilibrium with proper choice of parameters N, c_{th} and ϵ. The cost of the data analyst, which is the total expected payment at the equilibrium, goes to zero as the number of individuals goes to infinity. Since the privacy cost of an individual is always nonnegative, the total expected payment at an equilibrium of any mechanism is nonnegative due to individual rationality. Therefore, the designed mechanism asymptotically minimizes the cost for the data analyst to achieve any accuracy goal.

Recall that with the procured data $\boldsymbol{X}$, the data analyst learns the state W by performing hypothesis testing between the two hypotheses $H_0\colon W=0$ and $H_1\colon W=1$. An accuracy goal can be written as $p_e \leq p_e^{\max}$ for some $p_e^{\max}$, where p_e is the overall probability of error for hypothesis testing. We consider the maximum likelihood decision. The values for $N, c_{\text{th}}, \epsilon$ are chosen using the procedure below. The intuition is that we first fix the quality that the analyst expects to obtain from each participant and the types of individuals the analyst would like to collect data from, and then the accuracy goal can be met when the population size is large enough to make sure that there are enough participants.

Parameter Selection Procedure. Pick any ϵ such that $\epsilon \in (0, +\infty)$. Let

$$D(\epsilon) = \frac{1}{2}\ln\frac{(e^\epsilon+1)^2}{4(\theta e^\epsilon+1-\theta)((1-\theta)e^\epsilon+\theta)}, \quad n_e(\epsilon) = \frac{-\ln(\frac{1}{2}p_e^{\max})}{D(\epsilon)},$$
$$\rho(\epsilon) = \frac{1}{n_e(\epsilon)p_e^{\max}} + 2 + \sqrt{\frac{1}{(n_e(\epsilon))^2(p_e^{\max})^2} + \frac{2}{n_e(\epsilon)p_e^{\max}}}.$$

Then pick any integer N such that $N > \rho(\epsilon)n_e(\epsilon)$. For the selected N, let $p_{\text{th}}(N,\epsilon) = \rho(\epsilon)n_e(\epsilon)/N$, which is roughly the participation percentage, and then let $c_{\text{th}}(N,\epsilon) = \inf\{c\colon F_C(c) = p_{\text{th}}(N,\epsilon)\}$.

Recall that we assume F_C to be a continuous function, so the set $\{c\colon F_C(c) = p_{\text{th}}(N,\epsilon)\}$ is nonempty and thus $c_{\text{th}}(N,\epsilon) \geq 0$ is finite. An example of this parameter selection procedure (and the resulted upper bound on total expected payment) can be found in the full version [31].

Theorem 3. *For the mechanism $\boldsymbol{R}^{(N,\mathcal{P},c_{\text{th}},\epsilon)}$, consider the Bayesian Nash equilibrium $\boldsymbol{\sigma}$ given in Theorem 2. Given an accuracy goal $p_e \leq p_e^{\max}$, let $(N, c_{\text{th}}, \epsilon)$*

be chosen according to the parameter selection procedure above and the data analyst performs hypothesis testing using the maximum likelihood approach.

- *The decision function ψ has the following form:*

$$\psi(\boldsymbol{X}) = \begin{cases} 1 & \text{if } \sum_i \mathbb{1}_{\{X_i=1\}} \geq \sum_i \mathbb{1}_{\{X_i=0\}}, \\ 0 & \text{otherwise;} \end{cases} \tag{3}$$

- *The overall probability of error, p_e, meets the accuracy goal $p_e \leq p_e^{\max}$;*
- *The total expected payment is bounded as*

$$\mathbb{E}_{\boldsymbol{\sigma}}\left[\sum_{i=1}^{N} R_i^{(N,\mathcal{P},c_{\mathrm{th}},\epsilon)}(\boldsymbol{X})\right] \leq c_{\mathrm{th}}(\epsilon, N)\rho(\epsilon)n_e(\epsilon)\cdot(1+e^{-\epsilon}+\epsilon). \tag{4}$$

Since $\rho(\epsilon)$ and $n_e(\epsilon)$ are constants for given ϵ, and $c_{\mathrm{th}}(\epsilon, N)$ goes to 0 as $N \to \infty$, this total expected payment goes to zero, with the accuracy goal met, as $N \to \infty$.

The proof of Theorem 3 is presented in the full version [31]. Theorem 3 shows that choosing parameters according to the parameter selection procedure for the designed family of mechanisms not only meets the accuracy goal of the data analyst but is also cost-effective. The intuition is that as N becomes large, the requirement on the participation percentage becomes lower, which allows the mechanism to collect data from individuals with lower privacy costs and thus drives down the data analyst's cost. This suggests a way of constructing the asymptotically optimal sequence in Theorem 1: Fix an $\epsilon \in (0, +\infty)$, and then choose a sequence of mechanisms, each of which is designed for a different population size N and has parameter c_{th}, both of which are chosen according to the parameter selection procedure.

6 Conclusions

We considered incentive mechanisms for collecting private data from strategic, privacy-aware individuals, whose valuations of privacy are unknown. The data analyst is interested in learning an underlying state from the private data of individuals with minimum overall payment. We considered a local model of data privacy, where the data analyst is not necessarily trusted, and data subjects are endowed with the ability to control their own privacy, which frees the data analyst from the responsibility of privacy protection. We designed a family of payment mechanisms for the data analyst, which utilize negative payments to prevent individuals with high privacy valuations from reporting only noise and cut down the cost of the data analyst. In each designed mechanism, the individuals exhibit a threshold behavior at a Bayesian Nash equilibrium: only those with cost coefficients below some threshold participate, and they report data with certain quality guarantee, where the threshold and the quality guarantee are both parameters of the mechanism. With appropriate choices of parameters, the data analyst can fulfill any accuracy goal with diminishing cost at the equilibrium as the number of individuals grows to infinity.

Acknowledegments. This work was supported partially by NSF grant CNS-1618768.

References

1. Bassily, R., Smith, A.: Local, private, efficient protocols for succinct histograms. In: Proc. Ann. ACM Symp. Theory of Computing (STOC). pp. 127–135. Portland, OR (2015)
2. Cai, Y., Daskalakis, C., Papadimitriou, C.: Optimum statistical estimation with strategic data sources. In: Proc. Conf. Learning Theory (COLT). pp. 280–296. Paris, France (Jul 2015)
3. Chen, Y., Chong, S., Kash, I.A., Moran, T., Vadhan, S.: Truthful mechanisms for agents that value privacy. In: Proc. ACM Conf. Electronic Commerce (EC). pp. 215–232. Philadelphia, PA (2013)
4. Chen, Y., Sheffet, O., Vadhan, S.: Privacy games. In: Int. Conf. Web and Internet Economics (WINE). vol. 8877, pp. 371–385 (2014)
5. Cummings, R., Ligett, K., Roth, A., Wu, Z.S., Ziani, J.: Accuracy for sale: Aggregating data with a variance constraint. In: Proc. Conf. Innovations in Theoretical Computer Science (ITCS). pp. 317–324. Rehovot, Israel (2015)
6. Dasgupta, A., Ghosh, A.: Crowdsourced judgement elicitation with endogenous proficiency. In: Proc. Int. Conf. World Wide Web (WWW). pp. 319–330. Rio de Janeiro, Brazil (May 2013)
7. Duchi, J.C., Jordan, M.I., Wainwright, M.J.: Local privacy and minimax bounds: Sharp rates for probability estimation. In: Advances Neural Information Processing Systems (NIPS). pp. 1529–1537. Lake Tahoe, NV (Dec 2013)
8. Dwork, C.: Differential privacy. In: Proc. Int. Conf. Automata, Languages and Programming (ICALP). pp. 1–12. Venice, Italy (2006)
9. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Proc. Conf. Theory of Cryptography (TCC). pp. 265–284. New York, NY (2006)
10. Dwork, C., Roth, A.: The algorithmic foundations of differential privacy. Found. Trends Theor. Comput. Sci. 9(3–4), 211–407 (Aug 2014)
11. Erlingsson, Ú., Pihur, V., Korolova, A.: RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In: Proc. ACM SIGSAC Conf. Computer and Communication Security (CCS). pp. 1054–1067. Scottsdale, AZ (2014)
12. Fanti, G.C., Pihur, V., Erlingsson, Ú.: Building a RAPPOR with the unknown: Privacy-preserving learning of associations and data dictionaries. arXiv:1503.01214 [cs.CR] (2015)
13. Fleischer, L.K., Lyu, Y.: Approximately optimal auctions for selling privacy when costs are correlated with data. In: Proc. ACM Conf. Electronic Commerce (EC). pp. 568–585. Valencia, Spain (2012)
14. Ghosh, A., Ligett, K.: Privacy and coordination: Computing on databases with endogenous participation. In: Proc. ACM Conf. Electronic Commerce (EC). pp. 543–560. Philadelphia, PA (2013)
15. Ghosh, A., Ligett, K., Roth, A., Schoenebeck, G.: Buying private data without verification. In: Proc. ACM Conf. Economics and Computation (EC). pp. 931–948. Palo Alto, CA (2014)
16. Ghosh, A., Roth, A.: Selling privacy at auction. In: Proc. ACM Conf. Electronic Commerce (EC). pp. 199–208. San Jose, CA (2011)

17. Hsu, J., Khanna, S., Roth, A.: Distributed private heavy hitters. In: Proc. Int. Conf. Automata, Languages and Programming (ICALP). pp. 461–472. Warwick, UK (2012)
18. Kairouz, P., Oh, S., Viswanath, P.: Extremal mechanisms for local differential privacy. In: Advances Neural Information Processing Systems (NIPS). pp. 2879–2887. Montreal, Canada (Dec 2014)
19. Kasiviswanathan, S.P., Lee, H.K., Nissim, K., Raskhodnikova, S., Smith, A.: What can we learn privately? SIAM J. Comput. 40(3), 793–826 (May 2011)
20. Ligett, K., Roth, A.: Take it or leave it: Running a survey when privacy comes at a cost. In: Proc. Int. Workshop Internet and Network Economics (WINE). pp. 378–391. Liverpool, UK (2012)
21. Liu, Y., Chen, Y.: Learning to incentivize: Eliciting effort via output agreement. In: Proc. Int. Jt. Conf. Artificial Intelligence (IJCAI). New York, NY (Jul 2016)
22. McSherry, F., Talwar, K.: Mechanism design via differential privacy. In: Proc. Ann. IEEE Symp. Found. Comput. Sci. (FOCS). pp. 94–103. Providence, RI (2007)
23. Miller, N., Resnick, P., Zeckhauser, R.: Eliciting informative feedback: The peer-prediction method. In: Computing with Social Trust, pp. 185–212. Human–Computer Interaction Series, Springer London (2009)
24. Nissim, K., Vadhan, S., Xiao, D.: Redrawing the boundaries on purchasing data from privacy-sensitive individuals. In: Proc. Conf. Innovations in Theoretical Computer Science (ITCS). pp. 411–422. Princeton, NJ (2014)
25. Pai, M.M., Roth, A.: Privacy and mechanism design. SIGecom Exch. 12(1), 8–29 (Jun 2013)
26. Roth, A., Schoenebeck, G.: Conducting truthful surveys, cheaply. In: Proc. ACM Conf. Electronic Commerce (EC). pp. 826–843. Valencia, Spain (2012)
27. Shokri, R.: Privacy games: Optimal user-centric data obfuscation. In: Proc. Privacy Enhancing Technologies (PETS). pp. 299–315. Philadelphia, PA (2015)
28. Wang, W., Ying, L., Zhang, J.: On the relation between identifiability, differential privacy, and mutual-information privacy. In: Proc. Ann. Allerton Conf. Communication, Control and Computing. pp. 1086–1092. Monticello, IL (Sep 2014)
29. Wang, W., Ying, L., Zhang, J.: A game-theoretic approach to quality control for collecting privacy-preserving data. In: Proc. Ann. Allerton Conf. Communication, Control and Computing. pp. 474–479. Monticello, IL (Sep 2015)
30. Wang, W., Ying, L., Zhang, J.: A minimax distortion view of differentially private query release. In: Proc. Asilomar Conf. Signals, Systems, and Computers. pp. 1046–1050. Pacific Grove, CA (Nov 2015)
31. Wang, W., Ying, L., Zhang, J.: Buying data from privacy-aware individuals: The effect of negative payments. Tech. rep., Arizona State Univ., Tempe, AZ (2016)
32. Wang, W., Ying, L., Zhang, J.: The value of privacy: Strategic data subjects, incentive mechanisms and fundamental limits. In: Proc. Ann. ACM SIGMETRICS Conf. Antibes Juan-les-Pins, France (Jun 2016)
33. Warner, S.L.: Randomized response: A survey technique for eliminating evasive answer bias. J. Amer. Stat. Assoc. 60(309), 63–69 (Mar 1965)
34. Witkowski, J., Bachrach, Y., Key, P., Parkes, D.C.: Dwelling on the negative: Incentivizing effort in peer prediction. In: Proc. AAAI Conf. Human Computation and Crowdsourcing (HCOMP). Palm Springs, CA (Nov 2013)
35. Xiao, D.: Is privacy compatible with truthfulness? In: Proc. Conf. Innovations in Theoretical Computer Science (ITCS). pp. 67–86. Berkeley, CA (2013)