Title:	Mechanism Design and Differential Privacy
Name:	Kobbi Nisim[1], David Xiao[2]
Affil./Addr. 1:	Department of Computer Science, Ben-Gurion University, Israel
Affil./Addr. 2:	CNRS, Université Paris 7
Keywords:	Differential privacy; Mechanism design; Privacy-aware mechanism design; Purchasing privacy
SumOriWork:	2007; Frank McSherry and Kunal Talwar 2011; Arpita Ghosh and Aaron Roth 2012; Kobbi Nissim, Claudio Orlandi, and Rann Smorodinsky 2012; Lisa Fleischer and Yu-Han Lyu 2012; Katrina Ligett and Aaron Roth 2013; Yiling Chen, Stephen Chong, Ian A. Kash, Tal Moran, and Salil P. Vadhan 2014; Kobbi Nissim, Salil P. Vadhan, and David Xiao

Mechanism Design and Differential Privacy

Kobbi Nisim[1], David Xiao[2]

[1] Department of Computer Science, Ben-Gurion University, Israel
[2] CNRS, Université Paris 7

Years aud Authors of Summarized Original Work

2007; Frank McSherry and Kunal Talwar
2011; Arpita Ghosh and Aaron Roth
2012; Kobbi Nissim, Claudio Orlandi, and Rann Smorodinsky
2012; Lisa Fleischer and Yu-Han Lyu
2012; Katrina Ligett and Aaron Roth
2013; Yiling Chen, Stephen Chong, Ian A. Kash, Tal Moran, and Salil P. Vadhan
2014; Kobbi Nissim, Salil P. Vadhan, and David Xiao

Keywords

Differential privacy; Mechanism design; Privacy-aware mechanism design; Purchasing privacy

Problem Definition

Mechanism design and private data analysis both study the question of performing computations over data collected from individual agents while satisfying additional

restrictions. The focus in mechanism design is on performing computations that are compatible with the incentives of the individual agents, and the additional restrictions are towards motivating agents to participate in the computation (individual rationality) and towards having them report their true data (incentive compatibility). The focus in private data analysis is on performing computations that limit the information leaked by the output on each individual agent's sensitive data, and the additional restriction is on the influence each agent may have on the outcome distribution (differential privacy). We refer the reader to the sections on *algorithmic game theory* and on *differential privacy* for further details and motivation.

Incentives and privacy. In real-world settings, incentives influence how willing individuals are to part with their private data. For example, an agent may be willing to share her medical data with her doctor, because the utility from sharing is greater than the loss of utility from privacy concerns, while she would probably not be willing to share the same information with her accountant.

Furthermore, privacy concerns can also cause individuals to misbehave in otherwise incentive-compatible, individually-rational mechanisms. Consider for example a second-price auction: the optimal strategy in terms of payoff is to truthfully report valuations, but an agent may consider misreporting (or abstaining) because the outcome reveals the valuation of the second price agent, and the agent does not want to risk their valuation being revealed. In studies based on sensitive information, *e.g.* a medical study asking individuals to reveal whether they have syphilis, a typical individual with syphilis may be less likely to participate than a typical individual without the disease, thereby skewing the overall sample. The bias may be reduced by offering appropriate compensation to participating agents.

The Framework. Consider a setting with n individual agents, and let $x_i \in X$ be the private data of agent i for some type set X. Let $f : X^n \to Y$ be a function of the joint inputs of the agents $\mathbf{x} = (x_1, \ldots, x_n)$. Our goal is to build a mechanism M that computes $f(\mathbf{x})$ accurately and is compatible with incentives and privacy as we will now describe.

We first fix a function v that models the gain in utility that an agent derives from the *outcome* of the mechanism. We restrict our attention to a setting where this value can only depend on the agent's data and the outcome y of the mechanism:

$$v_i = v(x_i, y).$$

We also fix a function λ that models the loss in utility that an agent incurs because information about her private data is leaked by the outcome of the mechanism. Importantly, λ depends on the mechanism M, as the computation M performs determines the leakage. The loss can also depend on how much the agent values privacy, described by a parameter p_i (a real number in our modelling), on the actual data of all the individuals, on the outcome, as well as other parameters such as the strategy of the agent:

$$\lambda_i = \lambda(M, p_i, \mathbf{x}_{-i}, x_i, y, \ldots).$$

The overall utility that agent i derives from participating in the computation of M is

$$u_i = v_i - \lambda_i. \tag{1}$$

With this utility function in mind, our goal will be to construct truthful mechanisms M that compute f accurately. We note that in Equation 1 we typically think about both v_i and λ_i as positive quantities, but we do not exclude either of them being negative, so either quantity may result in a gain or a loss in utility.

We can now define the mechanism $M : X^n \times \mathbb{R}^n \to Y$ to be a randomized function taking as inputs the private inputs of the agents $\mathbf{x}$ and their privacy valuations $\mathbf{p}$ and returns a value in the set Y.

Modeling the privacy loss. In order to analyze specific mechanisms we will need to be able to control the privacy loss λ. Towards this end, we will need to assume that λ has some structure, and so we now discuss the assumptions we make and their justifications.

One view of privacy loss is to consider a framework of *sequential games*: an individual is not only participating in mechanism M, but she will also participate in other mechanisms $M', M'', \ldots$ in the future, and each participation will cause her to gain or lose in utility. Because her inputs to these functions may be correlated, revealing her private inputs in M may cause her to obtain less utility in the future. For example, an individual may hesitate to participate in a medical study because doing so might reveal she has a genetic predisposition to a certain disease, which may increase her insurance premiums in the future. This view is general and can formalize many of the concerns we typically associate with privacy: discrimination because of medical conditions, social ostracism, demographic profiling, etc.

The main drawback of this view is that it is difficult to know what the future mechanisms $M', M'', \ldots$ may be. However, if M is differentially private, then participating in M entails a guarantee that remains meaningful *even without knowing the future mechanisms*. To see this, we will use the following definition that is equivalent to the definition of ϵ-differential privacy [2]:

Definition 1 (Differential privacy). *A (randomized) mechanism $M : X^n \to Y$ is ϵ-differentially private if for all $\mathbf{x}, \mathbf{x}' \in X^n$ that differ on one entry, and for all $g : Y \to [0, \infty)$, it holds that*

$$\mathrm{Exp}[g(M(\mathbf{x}))] \leq e^{\epsilon} \cdot \mathrm{Exp}[g(M(\mathbf{x}'))],$$

where the expectation is over the randomness introduced by the mechanism M.

Note that $e^{\epsilon} \approx 1 + \epsilon$ for small ϵ, thus, if $g(y)$ models the expected utility of an individual tomorrow given that the result of $M(x) = y$ today, then by participating in a differentially private mechanism the individual's utility will change by at most ϵ.

Fact 1 *Let $g : Y \to [-1, 1]$. If M is ϵ-differential private then $\mathrm{Exp}[g(M(\mathbf{x}'))] - \mathrm{Exp}[g(M(\mathbf{x}))] \leq 2(e^{\epsilon} - 1) \approx 2\epsilon$ for all $\mathbf{x}, \mathbf{x}' \in X^n$ that differ on one entry.*

To see why this is true, let $g_-(y) = \mathbf{max}(0, -g(y))$ and $g_+(y) = \mathbf{max}(0, g(y))$. From Definition 1 and the bound on the outcome of g we get that $\mathrm{Exp}[g_+(M(\mathbf{x}'))] - \mathrm{Exp}[g_+(M(\mathbf{x}))] \leq (e^{\epsilon} - 1) \cdot \mathrm{Exp}[g_+(M(\mathbf{x}))] \leq e^{\epsilon} - 1$ and, similarly, $\mathrm{Exp}[g_-(M(\mathbf{x}))] - \mathrm{Exp}[g_-(M(\mathbf{x}'))] \leq e^{\epsilon} - 1$. As $g(y) = g_+(y) - g_-(y)$ we conclude that $\mathrm{Exp}[g(M(\mathbf{x}'))] - \mathrm{Exp}[g(M(\mathbf{x}))] \leq 2(e^{\epsilon} - 1)$.

With this in mind, we typically view λ as being "bounded by differential privacy" in the sense that if M is ϵ-differentially private, then $|\lambda_i| \leq p_i \cdot \epsilon$, where p_i (a positive real number) is an upper bound on the maximum value of $2|g(y)|$. In certain settings we make even more specific assumptions about λ_i, and these are discussed in the sequel.

Generic problems

We will discuss two generic problems for which key results will be given in the next section:

Privacy-aware mechanism design: Given an optimization problem $q : X^n \times Y \to \mathbb{R}$ construct a privacy-aware mechanism whose output $\hat{y}$ approximately maximizes $q(\mathbf{x}, \cdot)$. Using the terminology above, this corresponds to setting $f(\mathbf{x}) = \mathbf{argmax}_y q(\mathbf{x}, y)$, and the mechanism is said to compute $f()$ with accuracy α if (with high probability) $q(\mathbf{x}, f(\mathbf{x})) - q(\mathbf{x}, \hat{y}) \leq \alpha$. We mention two interesting instantiations of $q()$. When $q(\mathbf{x}, y) = \sum_i v(x_i, y)$ the problem is of maximizing social welfare. When x_i corresponds to how agent i values a digital good and $Y = \mathbb{R}^+$ is interpreted as a price for the good, setting $q(\mathbf{x}, y) = y \cdot |i : x_i \geq y|$ corresponds to maximizing the revenue from the good.

Purchasing privacy: Given a function $f : X^n \to Y$, construct a mechanism computing payments to agents for eliciting permission to use (some of) the entries of $\mathbf{x}$ in an approximation for $f(\mathbf{x})$. Here it is assumed that the agents cannot lie about their private values (but can misreport their privacy valuations). We will consider two variants of the problem. In the *insensitive value model* agents only care about the privacy of their private values $\mathbf{x}$. In the *sensitive value model* agents also care about the privacy of their privacy valuations $\mathbf{p}$, *e.g.* because there may be a correlation between x_i and p_i.

Basic differentially private mechanisms

We conclude this section with two differentially private mechanisms that are used in the constructions presented in the next section.

The Laplace Mechanism [2]. The Laplace distribution with parameter $1/\epsilon$, denoted $Lap(1/\epsilon)$ is a continuous probability distribution with zero mean and variance $2/\epsilon$. The probability density function of $Lap(1/\epsilon)$ is $h(z) = \frac{\epsilon}{2} e^{-\epsilon|z|}$. For $\Delta \geq 0$ we get $\Pr_{Z \sim Lap(1/\epsilon)}[|Z| > \Delta] = e^{-\epsilon\Delta}$.

Fact 2 *The mechanism M_{Lap} that on input $\mathbf{x} \in \{0,1\}^n$ outputs $y = \#\{i : x_i = 1\} + Z$ where $Z \sim Lap(1/\epsilon)$ is ϵ-differentially private. From the properties of the Laplace distribution we get that*

$$\Pr_{y \sim M_{Lap}(\mathbf{x})}[|y - \#\{i : x_i = 1\}| > \Delta] \leq e^{-\epsilon\Delta}.$$

The Exponential Mechanism [7]. Consider the optimization problem defined by $q : X^n \times Y \to \mathbb{R}$, where q satisfies $|q(\mathbf{x}, y) - q(\mathbf{x}', y)| \leq 1$ for all $y \in Y$ and all $\mathbf{x}, \mathbf{x}'$ that differ on one entry.

Fact 3 *The mechanism M_{Exp} that on input $\mathbf{x} \in X^n$ outputs $y \in Y$ chosen according to*

$$\Pr[y = t] = \frac{\exp\left(\frac{\epsilon}{2} q(\mathbf{x}, t)\right)}{\sum_{\ell \in Y} \exp\left(\frac{\epsilon}{2} q(\mathbf{x}, \ell)\right)} \tag{2}$$

is ϵ-differentially private. Moreover,

$$\Pr_{y \sim M_{Exp}(\mathbf{x})}[q(\mathbf{x}, y) \geq \mathrm{opt}(\mathbf{x}) - \Delta] \geq 1 - |Y| \cdot \exp\left(-\epsilon \cdot \Delta/2\right), \tag{3}$$

where $\mathrm{opt}(\mathbf{x}) = \mathbf{max} y \in Y(q(\mathbf{x}, y))$.

Notation. For two n-entry vectors $\mathbf{x}, \mathbf{x}'$ we write $\mathbf{x} \sim_i \mathbf{x}'$ to denote that they agree on all but the i-th entry. We write $\mathbf{x} \sim \mathbf{x}'$ if $\mathbf{x} \sim_i \mathbf{x}'$ for some i.

Key Results

The work of McSherry and Talwar [7] was first to realize a connection between differential privacy and mechanism design. They observed that (with bounded utility from the outcome) a mechanism that preserves ϵ-differential privacy is also ϵ-truthful, yielding ϵ-truthful mechanisms for approximately maximizing social welfare or revenue. Other works in this vein – using differential privacy but without incorporating the effect of privacy loss directly into the agent's utility function – include [11; 9; 5].

Privacy-aware mechanism design

The mechanisms of this section share the following setup assumptions:

Optimization problem. $q : X^n \times Y \to [0, n]$ and a utility function $U : X \times Y \to [0, 1]$.
Input. n players each having an input $x_i \in X$ and a privacy valuation p_i. The players may misreport x_i.
Output. The mechanism outputs an element $y \in Y$ approximately maximizing $q(\mathbf{x}, y)$.
Utility. Each player obtains utility $U(x_i, y) - \lambda_i$ where the assumptions on how the privacy loss λ_i behaves vary for the different mechanisms below and are detailed in their respective sections.
Accuracy. Let $\text{opt}(\mathbf{x}) = \mathbf{max}_{y \in Y}(q(\mathbf{x}, y))$. A mechanism is (Δ, δ)-accurate for all $\mathbf{x}$ it chooses $y \in Y$ such that $\Pr[\text{opt}(\mathbf{x}) - q(\mathbf{x}, y) \leq \Delta] \geq 1 - \delta$ where the probability is taken over the random coins of the mechanism. (One can also define accuracy in terms of $\text{opt}(\mathbf{x}) - \text{Exp}[q(\mathbf{x}, y)]$.)

Worst-case privacy model [8] In the worst-case privacy model the privacy loss of mechanism M is only assumed to be upper-bounded by the mechanisms privacy parameter, as in the discussion following Fact 1:

$$0 \leq \lambda_i \leq p_i \cdot \epsilon \quad \text{where} \quad \epsilon = \mathbf{max}_{x' \sim x, y \in Y} \ln \frac{\Pr[M(x) = y]}{\Pr[M(x') = y]}. \tag{4}$$

Nissim, Orlandi, and Smorodinsky [8] give a generic construction of privacy-aware mechanisms assuming an upperbound on the privacy loss as in Equation 4. The fact λ_i is only upper-bounded excludes the possibility of punishing misreporting via privacy loss (compare with algorithms 3, 4 below), and hence the generic construction resorts to a somewhat non-standard modeling from [9]. To illustrate the main components of the construction, we present a specific instantiation in the context of pricing a digital good, where such a non-standard modeling is not needed.

Pricing a digital good. An auctioneer selling a digital good wishes to design a single price mechanism that would (approximately) optimize her revenue. Every agent i has a valuation $x_i \in X = \{0, 0.01, 0.02, \ldots, 1\}$ for the good, and privacy preference p_i. Agents are asked to declare x_i to the mechanism, which chooses a price $y \in Y = \{0.01, 0.02, \ldots, 1\}$. Let x'_i be the report of agent i. If $x'_i < y$ then agent i does not pay nor receives the good and hence gains zero utility, *i.e.* $v_i = 0$. If $x'_i \geq y$ then agent i gets the good and pays y and hence gains in utility. We let this gain be $v_i = x'_i - y + 0.005$, where the additional 0.005 can be viewed as modeling a preference to receive the good (technically, this breaks the tie between the cases $x'_i = y$ and $x'_i = y - 1$). To summarize,

$$v(x_i, x'_i, y) = \begin{cases} x_i - y + 0.005 & \text{if } y < x'_i \\ 0 & \text{otherwise} \end{cases}$$

The privacy loss for agent i is from the information that may be potentially leaked on x'_i via the chosen price y. The auctioneer's optimal revenue is $\text{opt}(\mathbf{x}) = \mathbf{max}_{t\in Y}(t\cdot|\{i : x_i \geq t\}|)$ and the revenue she obtains when the mechanism chooses price y is $y\cdot|\{i : x'_i \geq y\}|$. The mechanism is presented in Algorithm 1.

Algorithm 1 (ApxOptRev)
Auxiliary input: privacy parameter ϵ, probability $0 < \eta < 1$.
Input: $\mathbf{x}' = (x'_1, \ldots, x'_n) \in X^n$.
ApxOptRev executes M_1 with probability $1-\eta$ and M_2 otherwise, where M_1, M_2 are:

M_1: Choose $y \in Y$ using the exponential mechanism, M_{Exp} (Fact 3), i.e.

$$\Pr[y = t] = \frac{\exp\left(\frac{\epsilon}{2}\cdot t\cdot|\{i : x'_i \geq t\}|\right)}{\sum_{\ell\in Y}\exp\left(\frac{\epsilon}{2}\cdot \ell\cdot|\{i : x'_i \geq \ell\}|\right)}.$$

M_2: Choose $y \in Y$ uniformly at random.

Agent utility. To analyze agent behavior, compare the utility of a misreporting agent to a truthful agent. (i) As Algorithm 1 is ϵ-differentially private, by our assumption on λ_i, by misreporting agent i may reduce her disutility due to information leakage by *at most* $p_i\cdot\epsilon$. (ii) Note that $v(x_i, x'_i, y) \leq v(x_i, x_i, y)$. Using this and Fact 1 we can bound the expected gain due to misreporting in M_1 as follows:

$$\text{Exp}_{y\sim M_1(\mathbf{x}'_{-i}, x'_i)}[v(x_i, x'_i, y)] - \text{Exp}_{y\sim M_1(\mathbf{x}'_{-i}, x_i)}[v(x_i, x_i, y)] \leq$$
$$\text{Exp}_{y\sim M_1(\mathbf{x}'_{-i}, x'_i)}[v(x_i, x'_i, y)] - \text{Exp}_{y\sim M_1(\mathbf{x}'_{-i}, x_i)}[v(x_i, x'_i, y)] \leq 2\cdot\epsilon.$$

(iii) On the other hand, in M_2, agent i loses at least $g = 0.01\cdot 0.005$ in utility whenever $x'_i \neq x'_i$, this is because y falls in the set $\{x_i{+}0.01, \ldots, x'_i\}$ with probability $x'_i - x_i \geq 0.01$ when $x_i < x'_i$, in which case she loses at least 0.005 in utility and, similarly, y falls in the set $\{x'_i, \ldots, x_i - 0.01\}$ with probability $x_i - x'_i \geq 0.01$ when $x_i < x'_i$, in which case she loses at least 0.005 in utility.

We hence get that agent i strictly prefers to report truthfully when

$$2\cdot\epsilon - \eta\cdot g + p_i\cdot\epsilon < 0. \tag{5}$$

Designer utility. Let m be the number of agents for which Equation 5 does not hold. We have $\text{opt}(\mathbf{x}') \geq \text{opt}(\mathbf{x}) - m$ and hence, using Fact 3, we get that

$$\Pr_{y\sim\text{ApxOptRev}(\mathbf{x}')}[y\cdot|\{i : x'_i \geq y\}| < \text{opt}(\mathbf{x}) - m - \Delta] \leq |Y|\cdot\exp(-\epsilon\Delta/2) + \eta$$
$$= 100\cdot\exp(-\epsilon\Delta/2) + \eta.$$

We omit from this short summary the discussion of how to choose the parameters ϵ and η (this choice directly affects m). One possibility is to assume the p_i has nice properties [8].

Per-outcome privacy model [1] In the output specific model, the privacy loss of mechanism M is evaluated on a per-output basis. Specifically, on output $y \in Y$ is is assumed that

$$|\lambda_i(\mathbf{x}, y)| \leq p_i\cdot F_i(\mathbf{x}, y) \text{ where } F_i(\mathbf{x}, y) = \mathbf{max}_{\mathbf{x}', \mathbf{x}''\sim_i\mathbf{x}} \ln\frac{\Pr[M(\mathbf{x}') = y]}{\Pr[M(\mathbf{x}'') = y]}. \tag{6}$$

To interpret Equation 6, consider an Bayesian adversary that has a prior belief μ on x_i and fix $\mathbf{x}_{-i}$. After seeing seeing $y = M(\mathbf{x}_{-i}, x_i)$ the Bayesian adversary updates her belief to μ'. For every event E defined over x_i, we get that

$$\mu'(E) = \mu(E|M(\mathbf{x}_{-i}, x_i) = y) = \mu(E) \cdot \frac{\Pr[M(\mathbf{x}_{-i}, x_i) = y|E]}{\Pr[M(\mathbf{x}_{-i}, x_i) = y]} \in \mu'(E) \cdot e^{\pm F_i(\mathbf{x},y)}.$$

This suggests that λ_i models harm that is "continuous" in the change in adversarial belief about i, in the sense that a small adversarial change in belief entails small harm. (Note, however, that this argument is restricted to adversarial beliefs on x_i *given* $\mathbf{x}_{-i}$.)

Comparison with worst-case privacy. Note that if M is ϵ-differentially private then $F_i(\mathbf{x}, y) \leq \epsilon$ for all $\mathbf{x}, y$. Equation 6 can hence be seen as a variant of Equation 4 where the fixed value ϵ is replaced with the output specific $F_i(\mathbf{x}, y)$. One advantage of such a per-outcome model is that the typical gain from misreporting is significantly smaller than ϵ. In fact, for all $\mathbf{x} \in X^n$ and $x'_i \in X$,

$$\left|\text{Exp}_{y\sim \text{M}(\mathbf{x})}\left[F_i(\mathbf{x}, y)\right] - \text{Exp}_{y\sim \text{M}(\mathbf{x}_{-i}, x'_i)}\left[F_i(\mathbf{x}, y)\right]\right| = O(\epsilon^2).$$

On the other hand, the modeled harm is somewhat weaker, as (by Fact 1) Equation 4 also captures harm that is not continuous in beliefs (such as decisions based on the belief crossing a certain threshold).

Assuming privacy loss is bounded as in Equation 6, Chen, Chong, Kash, Moran, and Vadhan [1] construct truthful mechanisms for an election between two candidates, facility location, and a VCG mechanism for public projects (the latter uses payments). Central to the constructions is the observation that F_i is large exactly when agent i has influence on the outcome of $M()$. To illustrate the main ideas in the construction we present here the two-candidate election mechanism.

Two-candidate election. Consider the setting of an election between two candidates. Every agent i has a preference $x_i \in X = \{A, B\}$, and privacy preference p_i. Agents are asked to declare x_i to the mechanism, which chooses an outcome $y \in Y = \{A, B\}$. The utility of agent i is then

$$v(x_i, y) = \begin{cases} 1 \text{ if } x = y \\ 0 \text{ otherwise} \end{cases}$$

The privacy loss for agent i is from the information that may be potentially leaked on her reported x'_i via the outcome y. The designer's goal is to (approximately) maximize the agents' social welfare (i.e., total utility from the outcome). The mechanism is presented in Algorithm 2.

Algorithm 2 (ApxMaj)
Auxiliary input: privacy parameter ϵ.
Input: $\mathbf{x}' = (x'_1, \ldots, x'_n) \in X^n$.
ApxMaj performs the following:

1. *Sample a value Z from Lap$(1/\epsilon)$.*
2. *Choose $y = A$ if $|\{j : x'_j = A\}| > |\{j : x'_j = B\}| + Z$ and $y = B$ otherwise.*

Agent utility. To analyze agent behavior, we compare the utility of a misreporting agent to a truthful agent. Notice that once the noise Z is fixed if agent i affects the outcome then her disutility from information leakage is at most $p_i \cdot \epsilon$ and her utility from the outcome decreases by 1. If agent i cannot affect the outcome then misreporting does not change either. We hence get that agent i strictly prefers to report truthfully when

$$p_i \cdot \epsilon < 1. \tag{7}$$

Note that by our analysis, Equation 7 implies *universal truthfulness* – agent i prefers to report truthfully for every choice of the noise Z. In contrast, Equation 5 only implies truthfulness in expectation.

Social welfare. Letting m be the number of agents for which Equation 7 does not hold, and using Fact 2, we get that Algorithm ApxMaj maximizes social welfare up to error $m + \frac{\log 1/\delta}{\epsilon}$ with probability $1 - \delta$. As in the previous section, we omit from this short summary the discussion of how to choose ϵ (this choice affects m and hence the accuracy of the mechanism).

Purchasing Privacy

The mechanisms of this section share the following setup assumptions, unless noted otherwise:

Input. n players each having a data bit $x_i \in \{0, 1\}$ and a privacy valuation $p_i > 0$. The players may misreport p_i but cannot misreport x_i. We will assume for convenience of notation that $p_1 \leq p_2 \leq \ldots \leq p_n$.

Intermediate outputs. The mechanism selects a subset of participating players $S \subset [n]$ and a scaling factor t, and a privacy parameter ϵ.

Output. The mechanism uses the Laplace mechanism to output an estimate $s = \frac{1}{t}\left(\sum_{i\in[S]} x_i + Z\right)$ where $Z \sim Lap(1/\epsilon)$, and payments v_i for $i \in [n]$.

Utility. Each player obtains utility $v_i - \lambda_i$ where the assumptions on how the privacy loss λ_i behaves vary for the different mechanisms below and are detailed in their respective sections.

Accuracy. A mechanism is α-accurate if $\Pr[|s - f(x)| \leq \alpha n] \geq 2/3$ where the probability is taken over the random coins of the mechanism.

We focus on designing mechanisms that approximate the sum function $f(x) = \sum_{i=1}^{n} x_i$ where each $x_i \in \{0, 1\}$, which has been the most widely studied function in this area. As one can see from the above setup assumptions, the crux of the mechanism design problem is in selecting the set S, choosing a privacy parameter ϵ, and computing payments for the players. We note that several of the works we describe below generalize beyond the setting we describe here (*i.e.* computing different f, fewer assumptions, etc). The following presentation was designed to give a unified overview (sacrificing some generality), but to preserve the essence both of the challenges posed by the problem of purchasing private data as well as each mechanism's idea in addressing the challenges.

Insensitive valuation model [4] In the insensitive valuation model the privacy loss λ_i of a mechanism M is assumed to be

$$\lambda_i = p_i \cdot \epsilon_i \text{ where } \epsilon_i = \mathbf{max}_{x,x' \sim_i x,p,s} \ln \frac{\Pr[M(x,p) = s]}{\Pr[M(x',p) = s]}. \tag{8}$$

It is named the insensitive valuation model because ϵ_i only measures the effect on privacy of changing player i's data bit, but not the effect of changing that player's privacy valuation.

Mechanisms. Two mechanisms are presented in the insensitive value model in [4], listed in Algorithms 3, 4. Algorithm 3 (FairQuery) is given a hard budget constraint and seeks to optimize accuracy under this constraint; Algorithm 4 (MinCostAuction) is given a target accuracy requirement and seeks to minimize payouts under these constraints.

Algorithm 3 (FairQuery)
Auxiliary input: budget constraint $B > 0$.

1. *Let $k \in [n]$ be the largest integer such that $p_k(n - k) \leq B/k$.*

2. *Select* $S = \{1, \ldots, k\}$ *and set* $\epsilon = \frac{1}{n-k}$. *Set the scaling factor* $t = 1$.
3. *Set payments* $v_i = 0$ *for all* $i > k$ *and* $v_i = \min\{\frac{B}{k}, p_{k+1}\epsilon\}$ *for all* $i \leq k$.

Algorithm 4 (MinCostAuction)
Auxiliary input: accuracy parameter $\alpha \in (0, 1)$.

1. *Set* $\alpha' = \frac{\alpha}{1/2+\ln 3}$ *and* $k = (1 - \alpha')n$.
2. *Select* $S = \{1, \ldots, k\}$ *and* $\epsilon = \frac{1}{n-k}$. *Set the scaling factor* $t = 1$.
3. *Set payments* $v_i = 0$ *for* $i > k$ *and* $v_i = p_{k+1}\epsilon$ *for all* $i \leq k$.

Guarantees. Algorithms 3 and 4 are individually rational and truthful. Furthermore, Algorithm 3 achieves the best possible accuracy (up to constant factors) for the class of *envy-free* and individually rational mechanisms, where the sum of payments to players does not exceed B. Algorithm 4 achieves the minimal payout (up to constant factors) for the class of *envy-free* and individually rational mechanisms that achieve α-accuracy.

Sensitive value model Ghosh and Roth [4] also defined the sensitive value model where λ_1 is as in Equation 8, except that ϵ_i is defined to equal

$$\mathbf{max}_{x,p,(x',p')\sim_i(x,p),s} \ln \frac{\Pr[M(x,p) = s]}{\Pr[M(x',p') = s]}. \tag{9}$$

Namely, we also measure the effect on the outcome distribution of the change in a single player's privacy valuation. It was shown in [4] and subsequent generalizations [10] that in this model and various generalizations where the privacy valuation itself is sensitive, it is impossible to build truthful, individually rational, and accurate mechanisms with worst case guarantees and making finite payments. To bypass these impossibility results, several relaxations were introduced.

Bayesian relaxation [3]. Fleischer and Lyu use the sensitive notion of privacy loss given in Equation 9. In order to bypass the impossibility results about sensitive values, they assume that the mechanism designer has knowledge of prior distributions P^0, P^1 for the privacy valuations. They assume that all players with data bit b have privacy valuation sampled independently according to P^b, namely that $p_i \overset{R}{\leftarrow} P^{x_i}$, independently for all i. Their mechanism is given in Algorithm 5.

Algorithm 5 (Bayesian mechanism from [3])
Auxiliary input: privacy parameter ϵ.

1. *Compute* $c = 1 - \frac{2}{\epsilon^2 n}$. *Compute* α_b *for* $b \in \{0, 1\}$ *such that* $\Pr_{p\overset{R}{\leftarrow}P^b}[p \leq \alpha_b] = c$.
2. *Set* S *be the set of players* i *such that* $p_i \leq \alpha_{x_i}$. *Set the scaling factor* $t = c$.
3. *For each player* $i \in S$, *pay* $\epsilon\alpha_{x_i}$. *Pay the other players* 0.

Algorithm 5 is truthful and individually rational. Assuming that the prior beliefs are correct, the mechanism is $O(\frac{1}{\epsilon n})$-accurate. The key use of knowledge of the priors is in accuracy: the probability of a player participating is c independent of its data bit.

Take-it-or-leave-it mechanisms [6]. Ligett and Roth put forward a setting where the privacy loss is decomposed into two parts

$$\lambda_i = \lambda_i^p + \lambda_i^x,$$

where λ_i^p is the privacy loss incurred by leaking information of whether or not an individual is selected to participate (*i.e.* whether individual i is in the set S), and where λ_i^x is the privacy loss incurred by leaking information about the actual data bit.

The interpretation is that a surveyor approaches an individual and offers them v_i to participate. The individual cannot avoid responding to this question and so unavoidably incurs a privacy loss λ_i^p without compensation. If he chooses to participate, then he loses an additional λ_i^x, but in this case he receives v_i in payment. While this is the framework we have been working in all along, up until now we have not distinguished between these two sources of privacy loss, rather considering only the overall loss. By explicitly separating them, [6] can make more precise statements about how incentives relate to each source of privacy loss.

In this model the participation decision of an individual is a function (only) of its privacy valuation, and so we define

$$\lambda_i^p = p_i \epsilon_i^p \text{ where } \epsilon_i^p = \mathbf{max}_{x,p,p' \sim_i p,s} \ln \frac{\Pr[M(x,p)=s]}{\Pr[M(x,p')=s]}. \tag{10}$$

We define $\lambda_i^x = p_i \epsilon_i^x$ where ϵ_i^x is as in the insensitive model, Equation 8. The mechanism is given in Algorithm 6.

Algorithm 6 (Take-it-or-leave-it mechanism [6])
Auxiliary input: accuracy parameter $\alpha \in (0,1)$, payment increment $\eta > 0$.

1. *Set $j = 1$ and $\epsilon = \alpha$.*
2. *Repeat the following:*
   a) *Set $E_j = 100(\log j + 1)/\alpha^2$ and $S_j = \emptyset$.*
   b) *For $i = 1$ until E_j:*
      i. *Sample without replacement $i \xleftarrow{R} [n]$.*
      ii. *Offer player i a payment of $(1+\eta)^j$.*
      iii. *If player i accepts, set $S_j = S_j \cup \{i\}$.*
   c) *Sample $\nu \xleftarrow{R} \Lambda(1/\epsilon)$. If $|S_j| + \nu \geq (1 - \alpha/8)E_j$, then break and output selected set $S = S_j$, privacy parameter ϵ, and normalizing factor $t = E_j$. For every $j' \leq j$, pay $(1+\eta)^{j'}$ to each player that accepted in round j' and pay 0 to all other players.*
   d) *Otherwise, increment j and continue.*

Algorithm 6 is α-accurate. It is not individually rational since players cannot avoid the take-it-or-leave-it offer, which leaks information about their privacy valuation that is not compensated. However, it is "one-sided truthful" in the sense that rational players will accept any offer v_i satisfying $v_i \geq \lambda_i^p - \lambda_i^x$. [6] also prove that for appropriately chosen η, the total payments made by Algorithm 6 not much more than that of the optimal envy-free mechanism making the same take-it-or-leave-it offers to every player.

Monotonic valuations [10]. Nissim, Vadhan, and Xiao [10] study a relaxation of sensitive values that they call *monotonic valuations*, where it is assumed that

$$\lambda_i(x,p) \leq p_i \cdot \epsilon_i^{\mathsf{mon}}(x,p) \text{ where } \epsilon_i^{\mathsf{mon}}(x,p) = \mathbf{max}_{(x',p') \sim_i^{\mathsf{mon}}(x,p),s} \ln \frac{\Pr[M(x,p)=s]}{\Pr[M(x',p')=s]}. \tag{11}$$

Here, $(x',p') \sim_i^{\mathsf{mon}} (x,p)$ denotes that $(x',p'),(x,p)$ are identical in all entries except the i'th entry and in the i'th entry it holds that either $x_i > x_i'$ and $p_i \geq p_i'$ both hold, or $x_i < x_i'$ and $p_i \leq p_i'$ both hold.

The intuition behind the definition is that for many natural settings, $x_i = 1$ is more sensitive than $x_i = 0$ (for example, if x_i represents whether an individual tested positive for syphilis), and it is therefore reasonable to restrict attention to the case where the privacy valuation when $x_i = 1$ is at least the privacy valuation when $x_i = 0$.

There are two other aspects in which this notion is unlike those used in the earlier works on purchasing privacy: (i) the definition may depend on the input, so the privacy loss may be smaller on some inputs than others, and (ii) we assume only an upper bound on the privacy loss, since $\epsilon_i^{\mathsf{mon}}$ does not say *which* information is leaked about player i, and so it may be that the harm done to player i is not as severe as $\epsilon_i^{\mathsf{mon}}$ would suggest. The mechanism is given in Algorithm 7.

Algorithm 7 (Mechanism for monotonic valuations [10])
Auxiliary inputs: budget constraint $B > 0$, privacy parameter $\epsilon > 0$.

1. *Set $\tau = \frac{B}{2\epsilon n}$.*
2. *Output selected set $S = \{i \mid p_i \leq \tau\}$, output privacy parameter ϵ, and scaling factor $t = 1$.*
3. *Pay B/n to players in S, pay 0 to others.*

Algorithm 7 is individually rational for all players and truthful for all players satisfying $p_i \leq \tau$. Assuming all players are rational, on inputs where there are h players having $p_i > \tau$, the mechanism is $(O(\frac{1}{\epsilon n}) + h)$-accurate. The accuracy guarantee holds regardless of how the players with $p_i > \tau$ report their privacy valuations.

Cross-References

Performance-Driven Clustering
Energy Minimization in VLSI Circuits

Recommended Reading

1. Yiling Chen, Stephen Chong, Ian A. Kash, Tal Moran, Salil P. Vadhan: Truthful mechanisms for agents that value privacy. EC 2013: 215-232.
2. Cynthia Dwork, Frank McSherry, Kobbi Nissim, Adam Smith: Calibrating Noise to Sensitivity in Private Data Analysis. TCC 2006: 265-284.
3. Lisa Fleischer, Yu-Han Lyu: Approximately optimal auctions for selling privacy when costs are correlated with data. EC 2012: 568-585.
4. Arpita Ghosh, Aaron Roth: Selling privacy at auction. EC 2011: 199-208.
5. Zhiyi Huang, Sampath Kannan: The Exponential Mechanism for Social Welfare: Private, Truthful, and Nearly Optimal. FOCS 2012: 140-149.
6. Katrina Ligett, Aaron Roth: Take It or Leave It: Running a Survey When Privacy Comes at a Cost. WINE 2012: 378-391.
7. Frank McSherry, Kunal Talwar: Mechanism Design via Differential Privacy. FOCS 2007: 94-103.
8. Kobbi Nissim, Claudio Orlandi, Rann Smorodinsky: Privacy-aware mechanism design. EC 2012: 774-789.
9. Kobbi Nissim, Rann Smorodinsky, Moshe Tennenholtz: Approximately optimal mechanism design via differential privacy. ITCS 2012: 203-213.
10. Kobbi Nissim, Salil P. Vadhan, David Xiao: Redrawing the boundaries on purchasing data from privacy-sensitive individuals. ITCS 2014: 411-422.
11. David Xiao: Is privacy compatible with truthfulness? ITCS 2013: 67-86.