Is privacy compatible with truthfulness?

David Xiao *

January 28, 2011

Abstract

We investigate the mainstream interpretation of differential privacy, which says that given a differentially private mechanism, people are likely to share their data truthfully because they are at little risk of revealing their own information. We argue that this interpretation is incomplete because people value their privacy, and so they should be *explicitly motivated* to reveal sensitive information. Using the notion of *truthfulness* from game theory, we show that in certain settings, existing differentially private mechanisms do not encourage participants to report their information truthfully.

On the positive side, we exhibit a transformation that takes truthful and efficient mechanisms and transform them into differentially private, truthful, and efficient mechanisms. Our transformation applies to games where the type space is small and the goal is to optimize a well-behaved quantity such as social welfare.

We also study what happens when an explicit numerical cost is assigned to the information leaked by a mechanism. We show that in this case, even differential privacy may not be strong enough of a notion to motivate people to participate truthfully.

*LIAFA, Université Paris 7 and Université Paris-Sud

1 Introduction

As our world becomes ever more digitized and connected, concerns about the privacy of our personal information have become increasingly pressing. With various organizations collecting, accessing, and storing individuals' data, our desire to fully take advantage of the data has come into stark tension with the equally important desire to keep information about various aspects of our lives private, including for example our medical histories and our demographic information.

Understanding and resolving this tension has been a long-standing goal in the statistics and data analysis literature [4, 13]. More recently, the theoretical computer science community has provided definitions and a rigorous treatment of this question [5, 6, 9, 3, 10], culminating in the definition and study of differential privacy [9, 6]. See [7] for a more complete overview of the area and further references.

Differential privacy guarantees the following: a (randomized) data release mechanism is ε-differentially private if, for any input database, any participant i in the database, and any possible output of the release mechanism s, the presence or absence of participant i causes at most a multiplicative e^{ε} change in the probability of the mechanism outputting s. The question is whether one can achieve privacy while at the same time preserving some notion of utility, since one could always build a trivially private (but not very useful) mechanism that outputs a constant value without looking at the input data. The advantage of working with the rigorous definition provided by differential privacy is that one can now show that certain tasks can be done with formal guarantees of utility and privacy, as indeed has been done for many problems [2, 3, 16, 18, 12, 20].

For a problem where one can achieve both meaningful utility and differential privacy, the differential privacy guarantee is interpreted as follows: since the output distribution of the mechanism is barely affected by the presence or absence of participant i in the above very precise and very strong sense, therefore the release mechanism leaks very little private information about participant i, and participant i should feel comfortable entering his/her data in the database.

In this paper we investigate this interpretation. The first question we pose is why does the participant want to join in the database at all? Does (s)he do so purely out of altruism for the organization gathering the data, or does (s)he do so out of self-interest? It seems like the first answer is unlikely, especially since the organization gathering the data is usually one that is held with some amount of suspicion, such as the government or commercial corporations.

Therefore we view the second answer, because individuals participate out of self-interest, as more realistic, and so the question of what motivates the participant to participate enters the realm of game theory. We will aim to understand the participants' actions by examining their self-interest using game-theoretic techniques, and seeing how privacy influences these questions.

Within game theory, this question falls naturally into the mechanism design framework. In mechanism design, one is given a game, which is specified by a number of players and possible types that they might have. Each type corresponds to some private information about the player (such as how much he values an item or where he is located), which he may choose to reveal truthfully or not. There is a space of all possible outputs for the game, and a global utility function that maps a setting of all the players' types plus an output of the game to a real number. There is an additional twist: the players' private information also gives them a *private* utility which depends on their type and the game's output. Each players' goal is to optimize their private utility, and this may conflict with optimizing the global utility, which might be interpreted as some kind of utility for society or for whoever set up the game.

Even without talking about privacy, designing mechanisms for games is a challenge. A mechanism is a (possibly randomized) function that takes a set of declared players' types and generates an output of the game and possibly payments to be made to or received from the players. One of the mechanism designer's foremost challenges is to come up with a mechanism that satisfies the following two conditions:

Truthfulness No rational player misreports their type because it would not increase their private utility.

Efficiency For every possible setting of the players' true types, the expected global utility of the output of the mechanism should be close to the maximum possible global utility.

One frequently studied global utility function is the *social welfare*, which is the sum of the private utilities of all the individuals. For games where the goal is to optimize social welfare, the VCG mechanism satisfies truthfulness and efficiency by using payments to align all the players' interests with the social welfare.

Observe that, in the game theoretic setting, truthfulness and participation are closely related notions: one can always extend a game to have a "$\perp$" type that represents non-participation, and have a player deviate to the "$\perp$" type to represent non-participation. Conversely, a player who reports a uniformly random type (independent of his true type) is in some sense choosing not to participate. In the rest of this paper, we will focus on truthfulness, with the understanding that non-participation is a specific kind of deviation from truthfulness.

We propose that what motivates people to participate in a database is that there is always a *game* associated with the database, and that participation in a database increases the utility of the players. For example this would be the case with a census, where the census data might be used to allocate government funds, or with commercial surveys, where there might be a reward for participating. We believe that having some utility attached to the database is unavoidable because, assuming that privacy has some inherent value (more on quantifying this later), the players would opt not to participate in a database at all without some explicit motivation. Therefore, the study of private data release mechanisms should be coupled with the study of the associated games.

More specifically, we study what happens when a database sanitization mechanism's output is simultaneously interpreted both as sensitive data to be made differentially private *as well as* the outcome to the game that is motivating players to participate in the database. We argue that in this case, truthfulness is a *necessary* property of the mechanism: indeed the original goal of preserving privacy is to motivate players to honestly submit their data to the database. Moreover, one could argue that if a player lies about their data then there is no reason for the database curator to try to help preserve that player's privacy.

We address two aspects of the relationship between private data release and their associated games. The first is whether one can construct mechanisms that are simultaneously differentially private, truthful, and efficient. We show the first mechanisms that achieve this by exhibiting a transformation from truthful and efficient mechanism into a private, truthful, and efficient mechanism for games where the type space is small and where the global utility is insensitive to any particular player's type and also depends only on the histogram of players' types.

In previous work on differential privacy (including the work in the first half of this paper), it is assumed that the utility derived from the mechanism is independent of the amount of information leaked by the mechanism about the players' private types. In the second part of the paper, we drop this assumption and assign a non-zero cost to the amount of information leaked by a mechanism, and we ask when does the cost of revealing information overwhelm the utility a player gets from the game? We will show that in such a setting, even differentially private mechanisms may not motivate individuals to be untruthful.

Previous work. There has already been a fruitful interaction between game theory and the study of differential privacy. The most closely related work is the mechanism of McSherry and Talwar [18], which achieves privacy, *approximate* truthfulness, and efficiency. Approximate truthfulness says that the amount any player can gain by announcing a false type is small (but possibly positive). Approximate truthfulness is in fact a consequence of differential privacy, which says that any one player's type can only affect the output distribution of the mechanism by a little. McSherry and Talwar [18] interpret approximate truthfulness to mean that the player might as well announce his type honestly, since he has little incentive to deviate. They also show that their mechanism achieves good efficiency (assuming the players behave truthfully).

Unfortunately, there are two drawbacks to this approach. First, Nissim et al. [21] observed that the MT mechanism suffers from the fact that *all* possible strategies of the player lead to approximately the same output distribution of the mechanism. If one accepts the MT interpretation that players are indifferent to small changes in their utility, then intuitively privacy may even motivate players to *lie* when interacting with the MT mechanism: players are indifferent to the small amount of utility from the game outcome that is lost by lying, but they would prefer to lie because that would reduce the amount of information leaked by the mechanism about their private type. (We will formalize this concern in Section 4, see also the discussion in Section 1.2.) Thus, it seems that approximate truthfulness cannot substitute for standard truthfulness when differential privacy is an explicit desired property.

Second, one might hope that the "approximate" part of the approximate truthfulness guarantee of McSherry and Talwar [18] might be just an artifact of the analysis, and that the mechanism might in fact be truthful. This was shown false, also in the work of Nissim et al. [21], who prove that the MT mechanism is *not* truthful for a simple game based on digital goods auctions.

Nissim et al. [21] go on to show that by combining the MT mechanism with a "Gap Mechanism" that incentivizes the players to report their types honestly, one can get a fully truthful mechanism. They apply this mechanism to give a truthful and approximately efficient mechanism for the k-facility problem on a line, which is a more general version of the 1-facility problem on a line that we will study in this paper. Unfortunately, their mechanism is not differentially private, because the Gap Mechanism relies on constraining the post-actions of the players, and this constraint reveals the types of the players in a very non-private way.

Ghosh and Roth [14] consider a question that is related but orthogonal to our work: they consider a game where each player has sensitive data that a database owner wants to gather. Their mechanism allows each player i to declare a pricing function $c_i : \mathbb{R} \to \mathbb{R}$ such that $c_i(\varepsilon)$ represents the minimal payment that player i would require to submit his data to a database that is then published after passing through an ε-differentially private sanitizer. Their mechanism uses these pricing functions to compute a value of ε and payment values that are paid out to each player such that enough of them are incentivized to participate to make the outcome of the sanitizer accurate. In their model, the players can only lie about the pricing function c_i, but not about the underlying sensitive data. Our model focuses on deviations resulting from players reporting false sensitive data.

Feigenbaum et al. [11] study whether it is possible to keep information private even from the database owner, *i.e.* before running sanitization. We do not study this problem here and treat the database owner as a trusted party. We note however that, using standard cryptographic assumptions and protocols, one can replace a trusted database owner by a secure multiparty computation among the individuals.

1.1 Private, truthful, and efficient (PTE) mechanisms

The MT mechanism is not necessarily truthful. Our first result revisits the question of whether or not the MT mechanism is truthful. We consider the well-studied 1-facility location on a line game, denoted LINE-1-FAC (see *e.g.* [19]). In the k-facility problem, each player is assigned a point in a metric space and the output of the game is a set of k points in the space (k facilities) so that the sum of the minimal distances between the players and the k facilities is minimized. In effect, it is the k-median clustering problem; however in the mechanism design world, one not only needs to come up with an efficient algorithm but a *truthful* algorithm so that players have no incentive to misreport their locations. It is known that there are no perfectly truthful and efficient mechanisms for this general problem, although approximately efficient mechanisms exist for $k = 1, 2$ [1, 17]. The LINE-1-FAC game is the restriction where $k = 1$ and forces the metric space to be the interval $[0, 1]$. In this case, the game *does* have a truthful and efficient mechanism, namely outputting the left-most median player's point [19] . We prove:

Theorem 1.1. *For LINE-1-FAC, the MT mechanism with any privacy $\varepsilon > 0$ is not truthful.*

We believe this theorem highlights the drawback of the MT mechanism in even sharper relief than the counter-example of [21], because the failure of the MT mechanism to be truthful stands in contrast to the fact that a truthful and efficient mechanism exists. The counter-example of [21], on the other hand, is a game for which it is easy to see that there exists *no* efficient and truthful mechanism (regardless of privacy).

Transforming a truthful and efficient mechanism into a private, truthful, and efficient mechanism. We give a transformation for a large class of games that converts truthful and efficient mechanisms into PTE mechanisms. As examples, we apply this transform to specific games, including the LINE-1-FAC game. In particular, this shows that the LINE-1-FAC game does have a PTE mechanism even though the MT mechanism fails to be truthful. To the best of our knowledge, these are the first PTE mechanisms exhibited for non-trivial games. We note here that our definition of privacy is the relaxed version of [8], which also allows a (negligible) additive difference in probability between two similar databases, in addition to the e^{ε} multiplicative difference.

We will consider games where the global utility is not too sensitive to any single player's type. The idea of the transformation is based on the ideas for privately releasing histogram data [9]. Suppose the type space of the game is a finite set of size q. For player inputs $\underline{t} = (t_1, \ldots, t_n)$ where t_i is the type of the i'th player, we will let $\underline{h}$ denote the histogram with q entries, one for each possible type, and where $h_j = |\{i \mid t_i = j\}|$, the number of players who have type j.

Suppose the utility function is $v(t, s)$ where t is a player type and s is an outcome of the game; suppose that v lies in the interval $[-1, 1]$. Suppose that the global utility function $w(\underline{t}, s)$ depends only on the histogram $\underline{h}$, and is insensitive: namely if $\underline{h}, \underline{h}'$ are close in ℓ_1 distance then for all s it holds that $|w(\underline{h}, s) - w(\underline{h}, s')|$ is small. For example, the social welfare function, which is just the sum of the individual players' utilities, is an example of such a global utility function. Without loss of generality, we can assume that all mechanisms for games with such global utility functions need only depend on the histogram of player types, rather than looking at the individual players' types. (We show in Appendix B.)

At a high level, our transformation from truthful and efficient to PTE works as follows.

1. Construct the histogram.
2. Add independent noise distributed according to the two-sided geometric distribution to each of the entries of the histogram. This is known to achieve differential privacy much as Laplace noise achieves privacy for continuous random variables [9, 15].
3. Run the original truthful and efficient mechanism on the perturbed histogram.

Some care must be taken that, in adding noise to the histogram, one does not generate negative entries into the histogram. Handling this will be the source of the small additive loss in privacy. This procedure gives the following theorem.

Theorem 1.2 (From truthful and efficient mechanisms to PTE mechanisms, informal. See Theorem 3.4.)**.** *Let $\varepsilon, \eta > 0$. Let G be a game where the type space is of size q and where the global utility function depends only on the histogram and is insensitive to individual players' types. Let M be a truthful and efficient mechanism. Then there exists a PTE mechanism M' for the game G with (ε, η)-differential privacy and whose approximation error is the approximation error of M plus $O(q \log(q/\eta)/\varepsilon)$.*

We will show that using a simple rounding procedure and then applying Theorem 1.2 to a mechansim for the discretized version of the LINE-1-FAC game, we can give a PTE mechanism for LINE-1-FAC.

Corollary 1.3. *For all $\varepsilon, \eta > 0$, there is a (ε, η)-differentially private, truthful, and $O((n^{1/2} \log(n/\eta)/\varepsilon)$ approximately efficient mechanism for LINE-1-FAC.*

The global utility in LINE-1-FAC is the social welfare, which takes value in $[-n, n]$. Thus, the approximation guarantee of M' is meaningful when the approximation error is $o(n)$, which is the case here. This stands in contrast to the MT mechanism, which as we noted is not truthful in this setting.

1.2 The value of privacy

In prior work, including the results outlined in the previous section, privacy and utility were orthogonal. One posited a utility measure and showed that one could achieve an approximately optimal solution that simultaneously satisfied differential privacy. However, we now propose that if one really believes that participants value their privacy, then this should explicitly be taken into account in their utility functions. Not only does this allow us to quantify *how much* participants value their privacy, it also allows us to model tradeoffs between utility and privacy. That is, participants may be willing to sacrifice some of the utility they reap from a game by reporting a false type and thereby protecting their privacy. Vice versa, participants may be willing to sacrifice their privacy if by doing so they gain a larger value from the outcome of the game.

We show that it is possible even for PTE mechanisms to leak too much information to achieve truthfulness in games where there exists a tradeoff between utility and privacy. We leave as an exciting and challenging open question whether there are games for which one can build truthful and efficient mechanisms when the players' utilities explicitly take into account the value of privacy.

Quantifying privacy The first task is to define a meaningful measure of how much information a mechanism leaks. It would be ideal to consider a worst-case notion of information cost, just as differential privacy is a worst-case notion, without supposing any distribution over the inputs. However, it seems problematic to define a worst-case notion that is compatible with the natural desiderata that the information cost should be 0 if player i reports a type that is independent of his honest type. This criterion implicitly requires that there is

a prior distribution over the player's type (since otherwise "independence" would be meaningless). Therefore we will work in the Bayesian setting and assume there is some prior distribution for each of the players' types. Furthermore, as is usually the case in the differential privacy literature, we assume that the adversary trying to learn about the players' types is computationally unbounded.

The most natural measure of information leakage in this setting is the *mutual information* between the type of a player and the output of the mechanism. Indeed, we will define the information cost as follows:

Definition 1.4. *The information cost of mechanism M to player i with prior distribution T_i and with other players' inputs fixed to $\underline{t}^{-i}$ is defined to be $\nu_i \cdot I(T_i; M(\underline{t}^{-i}, \sigma(T_i)))$ where ν_i is a non-negative weight and σ is player i's strategy (σ is the identity function if player i is truthful).*

Tradeoffs between the value of the game and the value of privacy. Given the definition of information cost in the previous section, we would like to explicitly incorporate it into the utility of the player. Let us use the word "game value" or simply "value" and the letter v to denote the benefit that the player extracts from the output of the game without taking into account privacy, and we will let overall utility, denoted by u, denote the expected game value minus the information cost. Formally:

$$u(T_i, \sigma, \underline{t}^{-i}) = \mathbb{E}_{t \leftarrow_R T_i}[v(t, M(\sigma(\underline{t}^{-i}, t)))] - \nu_i I(T_i; M(\underline{t}^{-i}, \sigma(T_i))) \tag{1.1}$$

Here, T_i is the prior distribution of player i's type, σ is player i's strategy, and $\underline{t}^{-i}$ is a fixing of the other players' types. The definition expresses the tradeoff nature of privacy, and while this simple linear tradeoff may not be the most accurate tradeoff for all settings, we believe it is an instructive starting point and reasonably captures the kind of tradeoffs that happen when real players consider their privacy.

Releasing histograms for the LINE-1-FAC game. Our negative result draws again on the LINE-1-FAC game. Corollary 1.3 says that a PTE mechanism for LINE-1-FAC exists, let us denote it by M (see Algorithm 3.12 for the definition of the mechanism). Roughly, M functions by rounding the positions of the players to discrete points, constructing a histogram of the players' rounded locations, perturbing the histogram, and finally outputting the median point of the perturbed histogram. Suppose now that, instead of outputting the median, the mechanism outputs "more than necessary" and publishes the entire perturbed histogram of the players' locations as well. Call this mechanism $\hat{M}$. We believe that $\hat{M}$, represents a plausible situation in the real world: an agency gathers data for a single explicit purpose, for example to decide where to build a pit stop on a highway so as best to serve residents living along the highway, but then publishes the data in a more general way so that it may be of use to other agencies that may want to use it in a different way.

$\hat{M}$ is actually PTE: truthfulness and efficiency remain because the facility location output is the same as M, while privacy also remains because the perturbation of the histogram was designed to be differentially private. However, we show in Theorem 4.2 that if player i's prior distribution of positions is uniform and we set the parameters so that the mechanism is $(\varepsilon, 2^{-\log^2 n})$-differentially private, then the mutual information of the histogram and player i's type is $\Omega(\nu_i \varepsilon^2)$. On the other hand, we show in Theorem 4.3 that the amount of utility that a player loses by ignoring his true type and declaring a fixed type, say 0, is at most $O(1/n^{1/4})$. This implies the following:

Theorem 1.5 (Informal, see Theorem 4.7)**.** *There is a mechanism for LINE-1-FAC which is private, truthful, and efficient if one considers only the value derived by the players, but which becomes untruthful when one takes into account the information cost if some player i gives weight $\nu_i \gg 1/n^{1/4}$ to the information cost.*

Typically, one thinks of $\nu_i = \Theta(1/\log n^{1/2})$ for this game, (see Remark 4.1 for a discussion). The above theorem therefore says that, unless one can be assured that the weight assigned to information cost by all the players is unusually small, one should not publish the entire histogram.

2 Preliminaries

We let $[q] = \{1, \ldots, q\}$. We identify sets S with the uniform distribution over the set. For a distribution X, we write $x \leftarrow_R X$ to denote a random sample from that distribution. For $x \in \mathbb{R}^n$, we let $\|\underline{x}\|_1 = \sum_{i=1}^n |x_i|$ denote

the ℓ_1 norm and $\|\underline{x}\|_\infty = \max_i |x_i|$ the ℓ_∞ norm. The mutual information between two random variables is $I(X;Y) = H(X) - H(X \mid Y)$, and the divergence between two variables that share the same support is $D(X\|Y) = \mathbb{E}_{x \leftarrow_{\mathrm{R}} X}(\log \Pr[X = x] - \log \Pr[Y = x])$. It is easy to check that $I(X;Y) = \mathbb{E}_{x \leftarrow_{\mathrm{R}} X} D(Y_x\|Y)$ where Y_x is the conditional distribution $(Y \mid X = x)$.

Mechanisms. A game with n players consists of the type space Q, a space S of possible outcomes of the game[1], a valuation function $v : Q \times S \to [-1, 1]$, and a welfare function $w : Q^n \times S \to \mathbb{R}$. When Q is finite, we let $|Q| = q$. For $i \in [n]$, for $t_1, \ldots, t_n$, we let $\underline{t}^{-i}$ denote the vector with $n-1$ entries given by $t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n$. We assume *w.l.o.g.* that the type space has a special $\perp$ type, and any player that has this special type is ignored. (Namely, an input with n players, k of whom have type $\perp$, is treated as an input to the game with $n-k$ players with those k players removed.)

A game has a *histogram-based* global utility function w if there is a function w' such that for all $\underline{t}, s$, it holds that $w(\underline{t}, s) = w'(\underline{h}(\underline{t}), s)$. We abuse notation and write $w(\underline{h}, s)$ when we mean $w(\underline{t}, s)$ where $\underline{h} = \underline{h}(\underline{t})$. A histogram-based global utility function is *insensitive* if $|\underline{h} - \underline{h}'|_1 \le k$ implies that for all s, $|w(\underline{h}, s) - w(\underline{h}', s)| \le k$.

The social welfare is $w(\underline{t}, s) = \sum_{i=1}^n v(t_i, s)$. Observe that it is histogram-based and insensitive.

A strategy σ for player i takes an input type t and produces an output distribution $\sigma(t)$. The player uses the strategy in a game by sampling $t' \leftarrow_{\mathrm{R}} \sigma(t)$ and reporting t' to the mechanism. The truthful strategy consists of reporting t itself.

A mechanism M is a (randomized) function that takes types $t_1, \ldots, t_n$ for all of the players and samples an outcome $(s, p_1, \ldots, p_n) \leftarrow_{\mathrm{R}} M(t_1, \ldots, t_n)$ where $s \in S$ and $p_i \in \mathbb{R}$. We say that M is *moneyless* if $p_1 = \ldots = p_n = 0$ for all inputs and random coins.

Definition 2.1 (Truthfulness)**.** 1. We say a mechanism M is *truthful* (in the worst case) if for each player $i \in [n]$, and for all $\underline{t}^{-i}, t_i, t_i'$, it holds that

$$\mathbb{E}_{(s,\underline{p}) \leftarrow_{\mathrm{R}} M(\underline{t})} M[v(t_i, s) - p_i] \ge \mathbb{E}_{(s,\underline{p}) \leftarrow_{\mathrm{R}} M(\underline{t}^{-i}, t_i')}[v(t_i, s) - p_i]$$

2. We say that M is *truthful for distributions* $T_1, \ldots, T_n$ if for all $i \in [n]$, all $\underline{t}^{-i}$, and all (possibly randomized) strategies $\sigma : Q \to Q$, it holds that

$$\mathbb{E}_{t_i \leftarrow_{\mathrm{R}} T_i, (s,\underline{p}) \leftarrow_{\mathrm{R}} M(\underline{t}^{-i}, t_i)}[v(t_i, s) - p_i] \ge \mathbb{E}_{t_i \leftarrow_{\mathrm{R}} T_i, (s,\underline{p}) \leftarrow_{\mathrm{R}} M(\underline{t}^{-i}, \sigma(t_i))}[v(t_i, s) - p_i]$$

Definition 2.2 (Efficiency)**.** We say a mechanism is $\delta(n)$-efficient if for all inputs $\underline{t}$ on n players, it holds that $\mathbb{E}_M[w(\underline{t}, M(\underline{t}))] \ge \max_{s \in S} w(\underline{t}, s; \omega) - \delta(n)$.

Observe that our definition of efficiency allows for an *additive* error. In contrast, most work in the mechanism design literature on approximate mechanisms deal with *multiplicative* error. However, additive error is more suitable when working with differential privacy.

Definition 2.3 (Differential privacy)**.** A mechanism M is (ε, η)-differentially private if for all $i \in [n]$, all $\underline{t}^{-i} \in Q^{n-1}$, all $t_i, t_i' \in Q$, and all $U \subseteq S \times \mathbb{R}^n$, it holds that $\Pr[M(\underline{t}^{-i}, t_i) \in U] \le e^{2\varepsilon} \Pr[M(\underline{t}^{-i}, t_i') \in U] + \eta$.

Typically we think of $\varepsilon > 0$ being a small constant and η as being $o(1)$, preferably negligible.

Two-sided geometric random variables and their truncations. The two-sided geometric distribution over the integers, centered at 0 with parameter ε is denoted $\mathcal{G}_\varepsilon$ and has probability mass function $f(x) = \frac{1-e^{-\varepsilon}}{1+e^{-\varepsilon}} e^{-\varepsilon|x|}$. This is the discrete version of the Laplace distribution. We will frequently write $\alpha = e^{-\varepsilon}$.

We will also use the following distribution $\mathcal{H}_{\varepsilon,\tau,q}$, which samples a vector of q variables as follows: sample $\underline{\zeta} \leftarrow_{\mathrm{R}} \mathcal{G}_\varepsilon^q$. If $\|\underline{\zeta}\|_\infty > \tau$, output 0, otherwise output $\underline{\zeta}$. We prove the following lemma in the appendix.

Lemma 2.4. *For all $i, j \in [q]$ and $U \subseteq \mathbb{Z}^q$, for $\underline{\zeta}'$ sampled from $\mathcal{H}_{\varepsilon,\tau,q}$ it holds that:*

$$\Pr[\underline{\zeta}' \in U] \le e^{2\varepsilon} \Pr[e_i - e_j + \underline{\zeta}' \in U] + \tfrac{2q\alpha^\tau}{1+\alpha}$$

[1] For simplicity we assume that the outcome space is the same for all n. Our main positive result Theorem 3.4 also holds if the outcome space depends on n with the additional property that solutions for n players can be truncated to solutions for $n' < n$ players.

It is straightforward to calculate that for $\zeta \leftarrow_{\mathrm{R}} \mathcal{G}_\varepsilon$:

$$\Pr[|\zeta| \geq \tau] = \frac{2\alpha^\tau}{1+\alpha} \tag{2.1}$$

Negative binomial distribution. The negative binomial distribution $\mathcal{NB}_{q,\varepsilon}$ is defined using the probability mass function

$$f(x) = \binom{x+q-1}{q-1} \cdot (1-e^{-\varepsilon})^q e^{-\varepsilon x}$$

We use the following fact about the collision probability of the negative binomial distribution.

Lemma 2.5. *For Y, Y' independently chosen from $\mathcal{NB}_{q,\varepsilon}$, it holds that*

$$\Pr[Y = Y'] \leq \frac{4\sinh(\varepsilon/2)}{\sqrt{q-1}}$$

The proof of this fact is somewhat non-trivial and we defer it to the appendix.
We also note the following facts:

Fact 2.6 (*e.g.* [22])**.** *1. The sum of q two-sided random geometric variables $\sum_{j=1}^{q} \zeta_j$, where each ζ_j is distributed according to $\mathcal{G}_\varepsilon$, is distributed identically to $Y - Y'$ where both Y, Y' are independent and identically distributed according to the $\mathcal{NB}_{q,\varepsilon}$.*

2. Suppose Y is distributed according to $\mathcal{NB}_{q,\varepsilon}$. Then $\Pr[Y \geq t] = \Pr[Z \leq q]$ where Z is a binomial random variable for an experiment with $q+t$ trials and success probability $(1-e^{-\varepsilon})$ for each trial.

3 PTE mechanisms

Definition 3.1. The LINE-1-FAC game is defined as follows. The player types are $t_i \in [0,1]$ and there is no public information. The output of mechanism is a point $s \in [0,1]$. The utility function is $v(t,s) = -|t-s|$ and the global utility is the social welfare: $w(\underline{t}, s) = \sum_{i\in[n]} v(t_i, s) = -\sum_{i\in[n]} |t_i - s|$.

We consider moneyless mechanisms for this game. It is well-known that the mechanism that outputs s to be the median (breaking ties say by picking the left median point) of the $\{t_1, \ldots, t_i\}$ is truthful and achieves optimal social welfare [19].

Theorem 1.1 (Restated)**.** *For LINE-1-FAC, the MT mechanism with any privacy $\varepsilon > 0$ is not truthful.*

Proof. For $n = 4$, set $t_1 = 0, t_2 = 0, t_3 = 2/3, t_4 = 1$. (This can easily be extended to an arbitrary even number of players n by placing $n/2$ players at 0 and $(n/2-1)$ players at 1, and one player at $2/3$.)

Claim 3.2. *For all $\varepsilon > 0$, if player 3 declares 1 then his utility under the MT mechanism is $-5/18$, while if he declares $2/3$ then his utility is strictly less than $-5/18$.*

Proof of Claim 3.2. The MT mechanism M_{MT} generates an output s according to the density function $f(s) = \frac{e^{\varepsilon w(\underline{t},s)}}{\int_0^1 e^{\varepsilon w(\underline{t},s)} ds}$. One can compute that the expected utility of player 3 if he reports 1 is $-\int_0^1 |2/3 - s| ds = -5/18$, since in this case there are exactly half the players at 0 and at 1 and so the social welfare function (and hence the mechanism's output) is uniform over $[0,1]$.

If player 3 is truthful, then the social welfare equals $w(\underline{t}, s) = -1 - s - |2/3 - s|$, and therefore the expected utility of player 3 is:

$$V \stackrel{def}{=} \mathbb{E}_{s\leftarrow_{\mathrm{R}} M_{\mathsf{MT}}(0,0,2/3,1)}[v_3(2/3, s)] = \frac{-\int_0^1 e^{-\varepsilon(1+s+|2/3-s|)}|2/3-s|ds}{\int_0^1 e^{-\varepsilon(1+s+|2/3-s|)}ds} = \frac{-\int_0^1 e^{-\varepsilon(s+|2/3-s|)}|2/3-s|ds}{\int_0^1 e^{-\varepsilon(s+|2/3-s|)}ds} \tag{3.1}$$

We claim that for all $\varepsilon > 0$, it holds that $V < -5/18$. This is equivalent to proving

$$0 < -V - 5/18 = \frac{\int_0^1 e^{-\varepsilon(s+|2/3-s|)} \cdot (|2/3-s| - 5/18)ds}{\int_0^1 e^{-\varepsilon(s+|2/3-s|)}ds}$$

Observe that the denominator is positive, so it suffices to prove that the numerator is positive for all $\varepsilon > 0$. Evaluating the integral we obtain that

$$\int_0^1 e^{-\varepsilon(s+|2/3-s|)} \cdot (|2/3-s| - 5/18)ds = e^{-2\varepsilon/3}\left(\frac{1}{27} + \frac{9-5\varepsilon-e^{-2\varepsilon/3}(\varepsilon+9)}{36\varepsilon^2}\right) \tag{3.2}$$

Since $e^{-2\varepsilon/3} > 0$ so we can multiply the LHS of Equation 3.2 by $36\varepsilon^2 e^{2\varepsilon/3}$ and simplify, and it suffices to prove that

$$\tfrac{4\varepsilon^2}{3} + 9 - 5\varepsilon - e^{-2\varepsilon/3}(\varepsilon+9) > 0 \tag{3.3}$$

For large $\varepsilon \gg 0$ the RHS of Equation 3.3 is dominated by the quadratic term. For instance it is easy to check that for all $\varepsilon > 3$, it holds that $e^{-2\varepsilon/3} < 1/7$ and so:

$$\begin{aligned}
\tfrac{4\varepsilon^2}{3} + 9 - 5\varepsilon - e^{-2\varepsilon/3}(\varepsilon+9) &> \tfrac{4\varepsilon^2}{3} + 9 - 5\varepsilon - \tfrac{\varepsilon}{7} - 9/7 && (3.4)\\
&> \tfrac{4\varepsilon^2}{3} - 5.2\varepsilon + 7.7 && (3.5)\\
&> 0 && (3.6)
\end{aligned}$$

where the final inequality can be deduced by the fact that the quadratic polynomial in Equation 3.5 has no real roots and is non-negative. Therefore we can restrict our attention to the case $\varepsilon \le 3$. Using the Taylor expansion of the exponential, we know that $e^{-2\varepsilon/3} \le 1 - \frac{2\varepsilon}{3} + \frac{2\varepsilon^2}{9} - \frac{4\varepsilon^3}{81} + \frac{2\varepsilon^4}{243}$, therefore Equation 3.3 can be rewritten as:

$$\begin{aligned}
\tfrac{4\varepsilon^2}{3} + 9 - 5\varepsilon - e^{-2\varepsilon/3}(\varepsilon+9) &\ge \tfrac{4\varepsilon^2}{3} + 9 - 5\varepsilon - (1 - \tfrac{2\varepsilon}{3} + \tfrac{2\varepsilon^2}{9} - \tfrac{4\varepsilon^3}{81} + \tfrac{2\varepsilon^4}{243})(\varepsilon+9)\\
&= \tfrac{4\varepsilon^2}{3} + 9 - 5\varepsilon - \varepsilon - 9 + \tfrac{2\varepsilon^2}{3} + 6\varepsilon - \tfrac{2\varepsilon^3}{9} - 2\varepsilon^2 + \tfrac{4\varepsilon^4}{81} + \tfrac{4\varepsilon^3}{9} - \tfrac{2\varepsilon^5}{243} - \tfrac{2\varepsilon^4}{27}\\
&= \tfrac{2\varepsilon^3}{9} - \tfrac{2\varepsilon^4}{81} - \tfrac{2\varepsilon^5}{243}\\
&> \varepsilon^3(\tfrac{2}{9} - \tfrac{2}{27} - \tfrac{2}{27})\\
&> 0
\end{aligned}$$

where in the last two lines we use the fact that $\varepsilon \le 3$. ∎

3.1 PTE achievable

We now exhibit a generic transformation that converts truthful and efficient mechanisms into PTE mechanisms.

Definition 3.3. For a game with type space of size q, the *histogram* associated with input $\underline{t}$, denoted by $\underline{h} = \underline{h}(\underline{t})$, is the q-dimensional vector where $h_j = |\{i \mid t_i = j\}|$.

We assume M satisfies the following: on input $\underline{t}$, M computes its output depending only on $\underline{h} = \underline{h}(\underline{t})$ (and never looks at the entries of $\underline{t}$ individually). In Appendix B we show that for a game where the global utility function is histogram-based, any mechanism can be converted into a mechanism satisfying this condition.

Theorem 3.4. *Let G be a game with type space of size q and whose global utility function is histogram-based and insensitive. Suppose G has a truthful and δ-efficient histogram mechanism. Suppose that $\delta(n) \in [0, n]$ is non-decreasing.*[2]

Then for all $\varepsilon, \eta > 0$, G also has a (ε, η)-differentially private, truthful, and δ'-efficient mechanism where $\delta'(n) = \delta(n + O(q\log(q/\eta)/\varepsilon)) + O(q\log(q/\eta)/\varepsilon)$.

Proof of Theorem 3.4. Let M be the truthful and δ-efficient mechanism. Set $\tau = O(\log(q/\eta)/\varepsilon)$ such that

$$\Pr[\|\underline{\zeta}\|_\infty \ge \tau] \le 2q\frac{e^{-\varepsilon\tau}}{1+e^{-\varepsilon}} \le \eta \tag{3.7}$$

In Algorithm 3.5 we construct a mechanism M' (using τ set as just mentioned) that is private, truthful, and δ'-approximately efficient. We prove that M' satisfies all three properties:

[2] One can assume that this condition always holds, since any δ not satisfying it can be made to satisfy it by artificially increasing δ at some points.

Input: types $t_1, \ldots, t_n \in [q]$. Auxiliary input: ε privacy parameter, $\tau = O(\log(q/\eta)/\varepsilon)$.

1. Sample $\underline{\zeta} \leftarrow_{\mathrm{R}} \mathcal{H}_{q,\varepsilon,\tau}$
2. Construct $\underline{h}' = \underline{h} + \underline{\zeta} + \tau \cdot \underline{1}$, where $\underline{1}$ is the all 1 vector.
3. Output $M(\underline{h}')$.

Algorithm 3.5. PTE mechanism based on a truthful and efficient mechanism M.

Truthfulness. Fix an input $\underline{t}$. We claim that for every choice of $\underline{\zeta}$, the mechanism is truthful. We run M on the histogram $\underline{h}' = \underline{h} + \underline{\zeta} + \tau \cdot \underline{1}$. This is a valid (*i.e.* non-negative) histogram since $\|\underline{\zeta}\|_\infty \leq \tau$. Since M is truthful on *all* histograms, it is also truthful on $\underline{h}'$. Observe that if a player benefits by deviating on the input $\underline{h}$, then that player would beneft by deviating on the input $\underline{h}'$ as well. But M is truthful on $\underline{h}'$, therefore there are no such deviations.

Efficiency. Let $\underline{\zeta}$ be sampled from $\mathcal{H}_{q,\varepsilon,\tau}$ and observe that satisfy $\|\underline{\zeta}\|_\infty \leq \tau$. For any input $\underline{t}$ define the modified histogram $\underline{h}' = \underline{h}(\underline{t}) + \underline{\zeta} + \tau \cdot \underline{1}$. Observe that $|\underline{h} - \underline{h}'|_1 \leq 2q\tau$. By the insensitivity of w, it holds that for all possible outcomes of the game s', it holds that $\left|w(\underline{h}', s') - w(\underline{t}, s')\right| \leq 2q\tau$. Therefore it holds that:

$$\begin{aligned}
\mathbb{E}_M[w(\underline{h}, M(\underline{h}'))] &\geq \mathbb{E}_{\underline{\zeta}, M}[w(\underline{h}', M(\underline{h}'))] - 2q\tau && (3.8)\\
&\geq \max_{s'} w(\underline{h}', s') - 2q\tau - \delta(n + 2q\tau) && (3.9)\\
&\geq w(\underline{h}', s_0) - 2q\tau - \delta(n + 2q\tau) && (3.10)\\
&\geq \max_{s} w(\underline{h}, s) - 4q\tau - \delta(n + 2q\tau) && (3.11)
\end{aligned}$$

In Equation 3.9 we use the efficiency of M and the fact that $\underline{h}'$ corresponds to a game with at most $n + 2q\tau$ players, as well as the assumption that $\delta(n) \leq \delta(n + 2q\tau)$. In Equation 3.11 s_0 denotes the outcome that maximizes $w(\underline{h}, s)$.

Applying Equation 3.11 to the expected utility of M', one obtains the following:

$$\mathbb{E}_{M'}[w(\underline{t}, M'(\underline{h}))] = \mathbb{E}_{\underline{\zeta} \leftarrow_{\mathrm{R}} \mathcal{H}_{q,\varepsilon,\tau}}[w(\underline{t}, M(\underline{h}'))] \geq \max_{s} w(\underline{t}, s) - \delta(n + 2q\tau) - 4q\tau$$

where $\underline{h}' = \underline{h}(\underline{t}) + \underline{\zeta} + \tau \cdot \underline{1}$.

Privacy. Observe that Algorithm 3.5 is just perturbing the histogram according to the distribution $\mathcal{H}_{\varepsilon,\tau,q}$. Therefore, from Lemma 2.4, we know that for any adjacent $\underline{h}, \underline{h}^*$, it holds that for all subsets $U \subseteq \mathbb{Z}^q$ that

$$\Pr[\underline{h} + \underline{\zeta}' \in U] \leq e^{2\varepsilon} \Pr[\underline{h}^* + \underline{\zeta}' \in U] + \tfrac{2qe^{-\varepsilon\tau}}{1+e^{-\varepsilon}}$$

Plugging in our choice of τ, this shows the mechanism is (ε, η)-differentially private. ■

3.2 LINE-1-FAC has a PTE mechanism

We show that although the MT mechanism is not truthful for LINE-1-FAC, there does exist a mechanism that is truthful, differentially private, and approximately efficient. This mechanism is given in Algorithm 3.12. The idea is to reduce LINE-1-FAC to a game D-L1F ("discrete one-facility on a line") with a small type space, and then to give a private, truthful, and efficient mechanism for D-L1F using Theorem 3.4.

Definition 3.6. The D-L1F$_\gamma$ game is defined as follows. Assume that $\gamma > 0$ is such that $q = 1/\gamma + 1$ is an integer. The player types are $t_i \in [q]$. Output of mechanism is $s \in [q]$. Utility function is $v(t, s) = -\gamma|t - s|$. Global utility is social welfare: $w(\underline{t}, s) = \sum_{i=1}^n v(t_i, s)$.

Input: histogram of player types $\underline{h}$. Let $n = \sum_{j=1}^{q} h_j$.

1. Output the minimal $s \geq 1$ such that $\sum_{j=1}^{s} h_j \geq \frac{n}{2}$.

Algorithm 3.7. Truthful and efficient mechanism for D-L1F

Theorem 3.8. *There is a truthful and perfectly efficient mechanism for D-L1F$_\gamma$.*

The mechanism is given in Algorithm 3.7. We may apply Theorem 3.4, we obtain:

Corollary 3.9. *For all $\varepsilon, \eta > 0$, D-L1F$_\gamma$ has a (ε, η)-differentially private, truthful, and δ-efficient mechanism for $\delta = O(\log(\frac{1}{\gamma\eta})/(\gamma\varepsilon))$.*

We now turn to proving Theorem 3.8

Proof of Theorem 3.8. The mechanism is listed in Algorithm 3.7.

Truthfulness. Since the the players are symmetric, it suffices just to consider the truthfulness of player 1. Fix $\underline{t}^{-1} = (t_2, \ldots, t_n)$. Let $\underline{h} = \underline{h}(\underline{t})$.

For all j it holds that $h_j \geq 0$. Furthermore, because player 1 is in column t_1, it holds that $h_{t_1} \geq 1$.

The mechanism's output is the minimal $s \geq 1$ such that $\sum_{j=1}^{s} h_j \geq n/2$. Let s be the output of the mechanism, and we consider what happens when t_1 declares some other value t^*. Let $\underline{h}^*$ be the histogram that is identical to $\underline{h}$ everywhere, except $h^*_{t_1} = h_{t_1} - 1 \geq 0$ and $h^*_{t^*} = h_{t^*} + 1$. Let s^* be the minimal s such that $\sum_{j=1}^{s^*} h^*_j \geq n/2$, namely the output of the mechanism on input $(\underline{t}^{-1}, t^*)$. We analyze the following cases, using the fact that both $\underline{h}, \underline{h}^*$ are non-negative:

1. $t_1 < s$. Because for all $s' < s$ it holds that $\sum_{j=1}^{s'} h^*_j \leq \sum_{j=1}^{s-1} h_j < n/2$, it follows that $s^* \geq s$. Since $t_1 < s$, this implies that $v(t_1, s^*) \leq v(t_1, s)$.
2. $t_1 = s$: in this case player 1's utility is 0, which cannot be improved (since for this game the utility is a non-positive number).
3. $t_1 > s$: Because it holds that $\sum_{j=1}^{s} h^*_j \geq \sum_{j=1}^{s} h_j \geq n/2$, therefore $s^* \leq s$. Since $t_1 > s$, this implies that $v(t_1, s^*) \leq v(t_1, s)$.

Therefore, regardless of the value of t_1, it holds that $v(t_1, s^*) \leq v(t_1, s)$, and therefore player 1 has no incentive to misreport his type.

Efficiency. Let s be the output of the mechanism. We prove the utility is greater for s than for all other s'.

Claim 3.10. *For any histogram $\underline{h}$ and $s = M(\underline{h})$, and for all $s' \in [q]$, it holds that $\sum_{j=1}^{q} h_j |j - s'| \geq \sum_{j=1}^{q} h_j |j - s|$.*

Recalling that $v(t, s) = -\gamma|t - s|$, this immediately implies that $w(\underline{t}, s') \leq w(\underline{t}, s)$ for all s'.

We now prove the claim. First consider $s' < s$. Split the summation $\sum_{j=1}^{q} h_j |j - s'|$ into three parts:

$$\sum_{j=1}^{q} h_j |j - s'| = \sum_{j \leq s'} h_j (s' - j) + \sum_{s' < j < s} h_j (j - s') + \sum_{s \leq j \leq q} h_j (j - s') \tag{3.12}$$

We will bound each of the three terms. Suppose first that $s' < s$, then the following hold:

$$\sum_{j \le s'} h_j(s'-j) = \sum_{j \le s'} h_j(s-j) - (s-s') \sum_{j \le s'} h_j \tag{3.13}$$

$$\sum_{s \le j} h_j(j-s') = \sum_{s \le j} h_j(j-s) - (s'-s) \sum_{s \le j} h_j \tag{3.14}$$

$$\begin{aligned} \sum_{s'<j<s} h_j(j-s') &= \sum_{s'<j<s} h_j(s-j) - \sum_{s'<j<s} h_j(s+s'-2j) \\ &\ge \sum_{s'<j<s} h_j(s-j) - (s-s') \sum_{s'<j<s} h_j \end{aligned} \tag{3.15}$$

The first two equalities follow by definition, while Equation 3.15 is justified by the inequality $s+s'-2j < s-s'$ and the fact that the h_j are non-negative.

Applying Equation 3.13 Equation 3.14, Equation 3.15 to Equation 3.12, we obtain

$$\sum_{j=1}^{q} h_j|j-s'| \ge (s-s') \left(\sum_{s \le j} h_j - \sum_{j<s} h_j \right) + \sum_{j=1}^{q} h_j|j-s| \tag{3.16}$$

Since $s - s' > 0$, $\sum_{j=1}^{q} h_j = n$, and, by the definition of s it holds that $\sum_{j<s} h_j < n/2$, it follows that the first term on the RHS of Equation 3.16 is positive. This implies the claim for the case $s' \le s$.

The case for $s' \ge s$ follows similarly, resulting in the inequality

$$\sum_{j=1}^{q} h_j|j-s'| \ge (s'-s) \left(\sum_{s<j} h_j - \sum_{j \le s} h_j \right) + \sum_{j=1}^{q} h_j|j-s|$$

Now, using the fact that $s' \ge s$ and by the definition of s, it holds that $\sum_{s<j} h_j \le n/2$, we can similarly conclude that the claim also holds in this case. ∎

3.3 Using D-L1F$_\gamma$ to give a PTE mechanism for LINE-1-FAC

Theorem 3.11. *For any $\gamma, \varepsilon, \eta > 0$, the mechanism of Algorithm 3.12 is (ε, η)-differentially private, truthful, and δ-efficient for the LINE-1-FAC game, where $\delta = n\gamma/2 + O(\frac{1}{\gamma\varepsilon} \log(\frac{1}{\gamma\eta}))$.*

As an example setting, pick $\gamma = n^{-1/2}$, which implies $\delta = O(n^{1/2} \log(n/\eta)/\varepsilon)$.

Proof of Theorem 3.11. (ε, η)-differential privacy follows immediately from Theorem 3.4. Efficiency is also straightforward, because the overall error is at most the discretization error, which is bounded by $\gamma/2$ for each player, plus the error from M', which is bounded by $O(\frac{1}{\gamma\varepsilon} \log(\frac{1}{\gamma\eta}))$.

Truthfulness. Truthfulness must be argued more carefully because the rounding process might cause unexpected problems. By symmetry, it suffices to consider player 1.

Fix $t_1, \underline{t}^{-1}$, and t^*. We will show that the player can gain no utility from declaring t^* when his actual type is t_1. Fix any choice of random coins $\underline{\zeta}$ used by M'. Let $\underline{h}' = \underline{h}(\underline{t}) + \underline{\zeta} + \tau \cdot \underline{1}$ and $\underline{h}^* = \underline{h}(\underline{t}^{-1}, t^*) + \underline{\zeta} + \tau \cdot \underline{1}$. Let $s = M(\underline{h}')$ and $s^* = M(\underline{h}^*)$. If $s = s^*$ then there is no advantage to declaring t^*. If $s^* \ne s$, then the truthfulness of M implies that $|\hat{t}_1 - s| \le |\hat{t}_1 - s^*| - 1$, since s, s^* are integers. The rounding process guarantees that for all $t \in [0,1]$, $|t - \gamma\hat{t}| \le \gamma/2$, therefore we may write

$$\begin{aligned} |t_1 - \gamma s| &\le \gamma|\hat{t}_1 - s| + \gamma/2 \\ &\le \gamma(|\hat{t}_1 - s^*| - 1) + \gamma/2 \\ &\le |t_1 - \gamma s^*| \end{aligned}$$

This again implies that there is no advantage to declaring t^*, and so since for every choice of $\underline{\zeta}$ the mechanism is truthful, it follows that the overall mechanism is truthful. ∎

Input: player types $t_1, \ldots, t_n \in [0,1]$. Assume for simplicity that $1/\gamma$ is an integer.

1. Discretize $[0,1]$ into $q = 1/\gamma + 1$ intervals: $[0, \gamma/2), [\gamma/2, (1+1/2)\gamma), \ldots, [(j-1/2)\gamma, (j+1/2)\gamma), \ldots, [(q-1/2)\gamma, 1]$.
2. Assign player i the proxy type $\hat{t}_i = j$ such that t_i falls into the j'th interval.
3. Let M denote the mechanism for $\mathsf{D\text{-}L1F}_\gamma$ given by Theorem 3.8. Let M' be the corresponding (ε, η)-differentially private, truthful, and efficient mechanism, given by Corollary 3.9. Run $M'(\hat{t}_1, \ldots, \hat{t}_n)$ to obtain $s \in [q]$.
4. Output γs.

Algorithm 3.12. (ε, η)-differentially private, truthful, and efficient mechanism for LINE-1-FAC.

4 The value of privacy

We now break away from the assumption that the players only care about the value derived from the outcome of a game, and we assign non-negative costs to the information leaked by mechanisms about the private types of the players. We use the definition of information cost from Definition 1.4. We recall the definition and attempt to justify some of its more unusual aspects.

Definition 1.4 (Restated)**.** *The information cost of mechanism M to player i with prior distribution T_i and with other players' inputs fixed to $\underline{t}^{-i}$ is defined to be $\nu_i \cdot I(T_i; M(\underline{t}^{-i}, \sigma(T_i)))$ where ν_i is a non-negative weight and σ is player i's strategy (σ is the identity function if player i is truthful).*

Privacy defined per strategy rather than per outcome. In standard game theory, utilities/costs are assigned to individual outcomes of games, rather than to strategies. However, this is impossible in our setting: as a thought experiment, consider the natural desirable property that a strategy that ignores the player's true type t should leak no information about t. This is impossible to achieve with a utility that depends only on the outcome, because the very notion of "ignoring the player's true type" only makes sense when considering the player's strategy on all types.

Privacy for player i depends on types of other players. One could consider the types of the other players as auxiliary information that the adversary may hold. In fact, this part of the definition could be relaxed (one could consider the randomness to be over all of the players' prior distributions). However, we prefer including fixing the other players' inputs as the notation is cleaner this way, and it models the fact that the adversary may know everything except player i's input.

Remark 4.1. Reasonable values of ν_i should be roughly $\Theta(1/H(T_i))$. This means the information cost is in some bounded range $[0, c]$ for some constant $c \leq 1$, which makes sense since the game value is in the bounded range $[-1, 1]$.

4.1 Releasing histograms

Perturbed histograms still leak information. Our first result shows that, if one tries to release histograms where each column is perturbed according to $\mathcal{G}_\varepsilon$, then a noticeable amount of information is still leaked.

Theorem 4.2. *Fix any q be a positive integer, $\underline{h} \in \mathbb{Z}^q$, $i \in [q]$ and J be uniformly distributed in $[q]$. Let $e_k \in \mathbb{Z}^q$ denote the k'th standard basis vector. Suppose $\underline{\zeta}$ is sampled according to the distribution $\mathcal{G}_\varepsilon^q$. Then:*

$$D(\underline{h} + e_i + \underline{\zeta} \quad \| \quad \underline{h} + e_J + \underline{\zeta}) \geq \frac{\tanh(\varepsilon/2)}{\ln(2)} \cdot (\varepsilon - 2\sinh(\varepsilon)/q)$$

Observe that since $I(\underline{h}+e_J \mid \underline{h}+e_J+\underline{\zeta}) = \mathbb{E}_{i\leftarrow_{\mathrm{R}} J} D(\underline{h}+e_i+\underline{\zeta} \| \underline{h}+e_J+\underline{\zeta})$, this also gives a lower bound on the information leaked by the histogram. Also, if $q > 2\sinh(\varepsilon)/\varepsilon$ then this is lower bounded by $\varepsilon \tanh(\varepsilon/2)/(2\ln 2)$. In particular, for $\varepsilon < 1$ and $q \geq 4$, the RHS is $\Omega(\varepsilon^2)$.

Proof of Theorem 4.2. We will let $\alpha = e^{-\varepsilon}$. Observe that the only randomness is from $\underline{\zeta}$ and J. Therefore we can develop the divergence expression as follows:

$$D(\underline{h}+e_i+\underline{\zeta} \quad \| \quad \underline{h}+e_J+\underline{\zeta}) = \mathbb{E}_{\underline{\zeta}}\left[\log \frac{\Pr_{\underline{\zeta}'}[\underline{\zeta}'=\underline{\zeta}]}{\Pr_{j\leftarrow_{\mathrm{R}} J, \underline{\zeta}'\leftarrow_{\mathrm{R}} \mathcal{G}_\varepsilon^q}[e_j+\underline{\zeta}'=e_i+\underline{\zeta}]}\right] \tag{4.1}$$

$$= -\tfrac{1}{\ln 2}\mathbb{E}_{\underline{\zeta}}\left[\ln \frac{\mathbb{E}_{j\leftarrow_{\mathrm{R}} J}[\Pr_{\underline{\zeta}'}[e_j+\underline{\zeta}'=e_i+\underline{\zeta}]]}{\Pr_{\underline{\zeta}'}[\underline{\zeta}'=\underline{\zeta}]}\right] \tag{4.2}$$

Since the coordinates of $\underline{\zeta}$ are independent, we may write

$$\Pr_{\underline{\zeta}'}[\underline{\zeta}'=\underline{\zeta}] = (\tfrac{1-\alpha}{1+\alpha})^q \alpha^{\sum_{k=1}^q |\zeta_k|} \tag{4.3}$$

$$\text{For } i\neq j,\ \Pr_{\underline{\zeta}'}[e_j+\underline{\zeta}'=e_i+\underline{\zeta}] = (\tfrac{1-\alpha}{1+\alpha})^q \alpha^{\sum_{k\in[q],k\neq i,j} |\zeta_k|+|\zeta_i+1|+|\zeta_j-1|} \tag{4.4}$$

Plugging this into Equation 4.2 we obtain that

$$D(\underline{h}+e_i+\underline{\zeta} \quad \| \quad \underline{h}+e_J+\underline{\zeta}) = -\frac{1}{\ln 2}\mathbb{E}_{\underline{\zeta}}\left[\ln \frac{1}{q}\left(1+\sum_{j\neq i}\alpha^{|\zeta_i+1|-|\zeta_i|+|\zeta_j-1|-|\zeta_j|}\right)\right]$$

Observe that for all $j\in[q]$, the expression $|\zeta_j-1|-|\zeta_j|$ takes value in $\{\pm 1\}$, and furthermore $\Pr[|\zeta_j-1|-|\zeta_j|=1] = \frac{1}{1+\alpha}$ and $\Pr[|\zeta_k-1|-|\zeta_k|=-1] = \frac{\alpha}{1+\alpha}$. The analogous fact holds for $|\zeta_i+1|-|\zeta_i|$. Therefore, since the coordinates of $\underline{\zeta}$ are independent, we may rename the variables to the following more convenient form:

$$-\mathbb{E}_{\underline{\zeta}}\left[\ln \frac{1}{q}\left(1+\sum_{j\neq i}\alpha^{|\zeta_i+1|-|\zeta_i|+|\zeta_j-1|-|\zeta_j|}\right)\right] = -\mathbb{E}_{\underline{x}}\left[\ln \frac{1}{q}\left(1+\sum_{j\neq i}\alpha^{x_i+x_j}\right)\right] \tag{4.5}$$

$$= -\mathbb{E}_{\underline{x}}\left[\ln \frac{\alpha^{x_i}}{q}\left(\alpha^{-x_i}+\sum_{j\neq i}\alpha^{x_j}\right)\right] \tag{4.6}$$

where each of the x_k is distributed according to $\Pr[x_k=1] = \frac{1}{1+\alpha}$ and $\Pr[x_k=-1] = \frac{\alpha}{1+\alpha}$.

$$-\mathbb{E}_{\underline{x}}\left[\ln \frac{\alpha^{x_i}}{q}\left(\alpha^{-x_i}+\sum_{j\neq i}\alpha^{x_j}\right)\right] = \frac{1-\alpha}{1+\alpha}\cdot\varepsilon - \mathbb{E}_{\underline{x}}\left[\ln \frac{1}{q}\left(\alpha^{-x_i}+\sum_{j\neq i}\alpha^{x_j}\right)\right] \tag{4.7}$$

$$= \frac{1-\alpha}{1+\alpha}\cdot\varepsilon - \mathbb{E}_{\underline{x}}\left[\ln\left(1+\frac{1}{q}\sum_{\ell=1}^{\infty}\frac{(\varepsilon x_i)^\ell}{\ell!}+\frac{1}{q}\sum_{j\neq i}\sum_{\ell=1}^{\infty}\frac{(-\varepsilon x_j)^\ell}{\ell!}\right)\right] \tag{4.8}$$

$$\geq \frac{1-\alpha}{1+\alpha}\cdot\varepsilon - \mathbb{E}_{\underline{x}}\left[\frac{1}{q}\sum_{\ell=1}^{\infty}\frac{(\varepsilon x_i)^\ell}{\ell!}+\frac{1}{q}\sum_{j\neq i}\sum_{\ell=1}^{\infty}\frac{(-\varepsilon x_j)^\ell}{\ell!}\right] \tag{4.9}$$

$$= \frac{1-\alpha}{1+\alpha}\cdot\varepsilon - (1-\tfrac{2}{q})\cdot\frac{1-\alpha}{1+\alpha}\cdot\sum_{\ell=1}^{\infty}\frac{(-\varepsilon)^{2\ell-1}}{(2\ell-1)!} - \sum_{\ell=1}^{\infty}\frac{\varepsilon^{2\ell}}{(2\ell)!} \tag{4.10}$$

$$= \frac{1-\alpha}{1+\alpha}\cdot\varepsilon + (1-\tfrac{2}{q})\tanh(\varepsilon/2)\cdot\sinh(\varepsilon) - \cosh(\varepsilon) + 1 \tag{4.11}$$

$$= \frac{1-\alpha}{1+\alpha}\cdot\varepsilon - \frac{2}{q}\cdot\tanh(\varepsilon/2)\sinh(\varepsilon) \tag{4.12}$$

$$= \tanh(\varepsilon/2)(\varepsilon - 2\sinh(\varepsilon)/q) \tag{4.13}$$

Equation 4.8 uses the Taylor series expansion of e^x. Equation 4.9 uses the inequality $\ln(1+x) \leq x$. Equation 4.10 uses the fact that $x_i^\ell = 1$ for even ℓ and $x_i^\ell = x_i$ for odd ℓ, and also the fact that $\mathbb{E}[x] = \frac{1-\alpha}{1+\alpha}$. Equation 4.11 uses the Taylor series expansions of $\sinh, \cosh$. Equation 4.12 and Equation 4.13 follow from simple algebra and standard hyperbolic trigonometric identities. ■

PTE mechanism for LINE-1-FAC is insensitive. Next we return to the PTE mechanism for LINE-1-FAC given by Algorithm 3.12 (one could also just consider the discrete version, Algorithm 3.7). Observe that (ε, η)-differential privacy implies that being untruthful can hurt the value of a player by at most $\varepsilon + \eta$. In fact, we show that for this particular mechanism, there exist inputs for which the loss is much smaller.

Theorem 4.3. *Fix n and $\varepsilon > 0$, an let M denote the mechanism for LINE-1-FAC, Algorithm 3.12 run with privacy (ε, η) and γ such that $1/\gamma + 1 = n^{1/2}$. Then there exists a setting of the remaining players' types $\underline{t}^{-1}$ such that for all values of t_1, t_1', it holds that:*

$$\mathbb{E}_M[v(t_1, M(\underline{t}^{-1}, t_1))] - \mathbb{E}_M[v(t_1, M(\underline{t}^{-1}, t_1'))] \leq O(\cosh(\varepsilon/2)/n^{1/4}) + \eta \tag{4.14}$$

Proof. To simplify our notation, suppose that the number of players is $n+1$ rather than n, and suppose $n^{1/2}$ is an integer. By our choice of γ, Algorithm 3.12 divides $[0,1]$ into $q = n^{1/2}$ intervals. The setting of player types $\underline{t}^{-1}$ is as follows: for $j = 0, \ldots, q-1$, place $n^{1/2}$ players at location γj.

Recall that M functions by rounding the players' locations to a histogram, generating noise distributed according to $\mathcal{H}_{\varepsilon,\tau,q}$ where $\tau = O(\log(q/\eta)/\varepsilon)$, adding it to the histogram, and then running the deterministic mechanism M of Algorithm 3.7. Notice that for our choice of $\underline{t}^{-1}$, there is no rounding necessary. Let us rewrite the LHS of Equation 4.14 as:

$$\mathbb{E}_{\underline{\zeta}}[v(t_1, M(\underline{h}(\underline{t}) + \underline{\zeta})) - v(t_1, M(\underline{h}(\underline{t}^{-1}, t_1') + \underline{\zeta}))] \tag{4.15}$$

The main observation is that for most choices of $\underline{\zeta}$, M gives the same output regardless of whether it was given t_1 or t_1'. To state this more formally, define

$$\underline{h}^{-1} = \underline{h}(\underline{t}^{-1}) + \underline{\zeta} + \tau \cdot \underline{1} \tag{4.16}$$

$$n' = (n+1) + \sum_{j=1}^{q} \zeta_j + q\tau \tag{4.17}$$

We are concerned with the event over choice of ζ that $M(\underline{h}(\underline{t}) + \underline{\zeta}) \neq M(\underline{h}(\underline{t}^{-1}, t_1') + \underline{\zeta})$. Call this event B. We claim that B is the contained in the union of the following two events B_1 and B_2:

1. Let B_1 denote the event over the choice of $\underline{\zeta}$ that $\|\underline{\zeta}\|_\infty > \tau$.

2. Let B_2 denote the event over choice of $\underline{\zeta}$ that $\|\underline{\zeta}\|_\infty \leq \tau$ and there exists $k \in [q]$ such that $\sum_{j=1}^{k} (\underline{h}^{-1})_j = \lceil n'/2 - 1 \rceil$.

It follows from the definition of the median computed by Algorithm 3.7 that the only time a single player can affect the median is when B occurs: if neither B_1 nor B_2 occurs, then regardless of what player 1 declares, the median remains fixed. Therefore we bound $\Pr[B] \leq \Pr[B_1] + \Pr[B_2]$.

It follows from Equation 2.1 and the choice of $\tau = O(\log(n/\eta)/\varepsilon)$ that $\Pr[B_1] = q \Pr[|\underline{\zeta}| > \tau] \leq \frac{2qe^{-\varepsilon(\tau+1)}}{1+e^{-\varepsilon}} \leq \eta$.

Claim 4.4. $\Pr[B_2] \leq O(\cosh(\varepsilon/2)/n^{1/4})$

From this claim, we conclude that $\Pr[B] \leq \eta + O(\cosh(\varepsilon/2)/n^{1/4})$. Since we argued above that

$$\mathbb{E}_{\underline{\zeta}}[v(t_1, M(\underline{h}(\underline{t}) + \underline{\zeta})) - v(t_1, M(\underline{h}(\underline{t}^{-1}, t_1') + \underline{\zeta}))] \leq \Pr_{\underline{\zeta}}[B] \tag{4.18}$$

this implies the theorem. ■

Proof of Claim 4.4. Observe that

$$\Pr_{\underline{\zeta}\leftarrow_{\mathrm{R}}\mathcal{H}_{\varepsilon,q,\tau}}[B_2] = \Pr_{\underline{\zeta}\leftarrow_{\mathrm{R}}\mathcal{H}_{\varepsilon,q,\tau}}\left[\|\underline{\zeta}\|_\infty \le \tau \quad \wedge \quad \exists k,\ \sum_{j=1}^{k}(\underline{h}^{-1})_j = \lceil n'/2-1\rceil\right] \tag{4.19}$$

$$= \Pr_{\underline{\zeta}\leftarrow_{\mathrm{R}}\mathcal{G}_\varepsilon^q}\left[\|\underline{\zeta}\|_\infty \le \tau \quad \wedge \quad \exists k,\sum_{j=1}^{k}(\underline{h}^{-1})_j = \lceil n'/2-1\rceil\right] \tag{4.20}$$

$$\le \Pr_{\underline{\zeta}\leftarrow_{\mathrm{R}}\mathcal{G}_\varepsilon^q}\left[\exists k,\ \sum_{j=1}^{k}(\underline{h}^{-1})_j = \lceil n'/2-1\rceil\right] \tag{4.21}$$

$$\le \sum_{k=1}^{q}\Pr_{\underline{\zeta}\leftarrow_{\mathrm{R}}\mathcal{G}_\varepsilon^q}\left[\sum_{j=1}^{k}(\underline{h}^{-1})_j = \lceil n'/2-1\rceil\right] \tag{4.22}$$

where Equation 4.20 holds because by definition the distribution of $\underline{\zeta}\leftarrow_{\mathrm{R}}\mathcal{H}_{\varepsilon,q,\tau}$ is exactly the same as $\mathcal{G}_\varepsilon^q$ under the condition $|\underline{\zeta}\|_\infty \le \tau$.

Observe that, by the definition of $\underline{h}^{-1}$ and n', we may rewrite Equation 4.22 to obtain:

$$\Pr_{\underline{\zeta}}[B_2] \le \sum_{k=1}^{q}\ \Pr_{\underline{\zeta}}\left[\sum_{j=1}^{k}(\underline{h}(\underline{t}^{-1})_j + \zeta_j) = -b + \sum_{j=k+1}^{q}(\underline{h}(\underline{t}^{-1})_j + \zeta_j) + (q-2k)\tau\right] \tag{4.23}$$

where $b=1$ if n' is even and $b=0$ if n' is odd. Let us consider $b=0$, for the other case follows by the same argument. By construction there are $n^{1/2}$ players in each of the first $n^{1/2}$ columns of the histogram, it follows that $\sum_{j=1}^{k}\underline{h}(\underline{t}^{-1})_j = kn^{1/2}$. Therefore the RHS of Equation 4.23 equals

$$\Pr_{\underline{\zeta}}\left[\sum_{j=1}^{k}\zeta_j = \sum_{j=k+1}^{q}\zeta_j + (q-2k)(n^{1/2}+\tau)\right]$$

Since $\mathcal{G}_\varepsilon$ is symmetric therefore ζ_j is distributed identically as $-\zeta_j$, and combined with the above we may deduce:

$$\Pr_{\underline{\zeta}}[B_2] \le \sum_{k=1}^{q}\Pr_{\underline{\zeta}}\left[\sum_{j=1}^{q}\zeta_j = |2k-q|(n^{1/2}+\tau)\right] \tag{4.24}$$

We divide the summation into two cases, first when k is far from $q/2$ and second when k is close to $q/2$. In the following, let $\alpha = e^{-\varepsilon}$.

Lemma 4.5. *For all k such that $|2k-q| \ge \frac{(1+\alpha)}{1-\alpha}$, it holds that*

$$\Pr_{\underline{\zeta}}\left[\sum_{j=1}^{q}\zeta_j = |2k-q|(n^{1/2}+\tau)\right] \le e^{-\sqrt{n}/8}$$

Lemma 4.6. *For all k (in particular $|2k-q| \le \frac{(1+\alpha)}{1-\alpha}$), it holds that*

$$\Pr_{\underline{\zeta}}\left[\sum_{j=1}^{q}\zeta_j = |2k-q|(n^{1/2}+\tau)\right] \le \frac{4\sinh(\varepsilon/2)}{\sqrt{n^{1/2}-1}} \tag{4.25}$$

Combining the two lemmas with Equation 4.24 to obtain:

$$\Pr_{\underline{\zeta}}[B_2] \leq qe^{-\sqrt{n}/8} + \frac{1+\alpha}{1-\alpha} \cdot \frac{4\sinh(\varepsilon/2)}{\sqrt{n^{1/2}-1}} \tag{4.26}$$

$$=\leq qe^{-\sqrt{n}/8} + \coth(\varepsilon/2) \cdot \frac{4\sinh(\varepsilon/2)}{\sqrt{n^{1/2}-1}} \tag{4.27}$$

$$=\leq qe^{-\sqrt{n}/8} + \cdot \frac{4\cosh(\varepsilon/2)}{\sqrt{n^{1/2}-1}} \tag{4.28}$$

$$\tag{4.29}$$

which, for large n, is $O(\cosh(\varepsilon/2)/n^{1/4})$. ■

We now turn to proving Lemma 4.5 and Lemma 4.6. The key to proving both these lemmas is the characterization of the sum of q two-sided geometric random variables given in Fact 2.6: $\sum_{j=1}^{q} \zeta_j$ is distributed identically to $Y - Y'$ where Y, Y' are independent $\mathcal{NB}_{q,\varepsilon}$ variables.

Proof of Lemma 4.5. Since $|2k - q| \geq \frac{1+\alpha}{1-\alpha}$ and $n^{1/2} + \tau \geq q$, it suffices to upper bound

$$\Pr_{\underline{\zeta}}\left[\sum_{j=1}^{q} \zeta_j \geq \frac{(1+\alpha)q}{1-\alpha}\right] \tag{4.30}$$

By Fact 2.6 it holds that $\sum_{j=1}^{q} \zeta_j$ is distributed identically to $Y - Y'$ as stated in Fact 2.6. Furthermore, since both Y, Y' are non-negative, $Y - Y' \geq \frac{(1+\alpha)q}{1-\alpha}$ implies that $Y \geq \frac{(1+\alpha)q}{1-\alpha}$. Therefore to bound Equation 4.30, it suffices to bound $\Pr[Y \geq \frac{(1+\alpha)q}{1-\alpha}]$. We apply the second point of fact Fact 2.6, which says that this probability is equal to the probability $\Pr[Z \leq q]$ where Z is a binomial random variable with $\frac{(1+\alpha)q}{1-\alpha} + q = \frac{2q}{1-\alpha}$ trials and success probability $1 - \alpha$. This means that $\mathbb{E}[Z] = 2q$. We can apply the Hoeffding bound for binomial variables to conclude that:

$$\Pr[Z \leq q] \leq e^{-(1/2)^2(1-\alpha)(2q/(1-\alpha))/4} = e^{-q/8} \tag{4.31}$$

■

Proof of Lemma 4.6. We will apply Lemma 2.5. Since $\sum_{j=1}^{q} \zeta_j$ is distributed like $Y - Y'$ where Y, Y' are independent $\mathcal{NB}_{q,\varepsilon}$ variables, the lemma is equivalent to bounding

$$\Pr[Y - Y' = |2k - q|(n^{1/2} + \tau)]$$

Observe that, by an application of Cauchy-Schwarz on the probability vector of $\mathcal{NB}_{q,\varepsilon}$, we have that, for all integers c

$$\Pr[Y - Y' = c] \leq \Pr[Y - Y' = 0]$$

Therefore applying Lemma 2.5 implies the lemma. ■

Releasing the histogram can compromise truthfulness. Given Theorem 4.3 and Theorem 4.2, we show that releasing the histogram produced in the PTE mechanism for the LINE-1-FAC game can break truthfulness. Let M be the mechanism of Algorithm 3.12 with privacy (ε, η). Recall that, essentially, this mechanism generates perturbations for a histogram and then outputs the median point of the perturbed histogram. We now modify this mechanism so that in addition to the output facility location, it also outputs the perturbed histogram $\underline{h}' = \underline{h}(\underline{t}) + \underline{\zeta} + \tau \cdot \underline{1}$. (To simplify the analysis, suppose it outputs the histogram with perturbation $\underline{\zeta}$, even if $\|\underline{\zeta}\|_\infty > \tau$.) Call this modified mechanism $\hat{M}$, and notice that $\hat{M}$ remains PTE: privacy holds because the perturbed histogram is (ε, η)-differentially private, while truthfulness and efficiency remain (with respect just to the game value and without taking into account the information cost) because the facility location output is the same as what M would have output.

However, $\hat{M}$ is *not* truthful when one takes into account the information cost:

Theorem 4.7 (Formal statement of Theorem 1.5)**.** *Fix any $\varepsilon > 0$ a constant, $\eta = o(n^{-1/4})$. Suppose that there exists a player i such that $\nu_i = \omega(1/n^{-1/4})$. Let T_i be uniformly distributed over the points $0, \gamma, 2\gamma, \ldots, 1$ where $1/\gamma + 1 = n^{1/2}$. Then $\hat{M}$ is not truthful with respect to the distribution T_i and the tradeoff utility function:*

$$u(T_i, \underline{t}^{-i}, \sigma) = \mathbb{E}[v(T_i, \hat{M}(\underline{t}^{-i}, \sigma(T_i)))] - \nu_i I(T_i; \hat{M}(\underline{t}^{-i}, \sigma(T_i))) \tag{4.32}$$

where v is the function computes the distance between the player and the facility contained in the output of $\hat{M}$.

Proof. A uniformly random location for player i adds 1 to a uniformly random column of the histogram of player locations. Therefore, applying Theorem 4.2 and the fact that ε is constant while q grows with n, we may write:

$$\forall \underline{t}^{-i},\ I(T_i; \hat{M}(\underline{t}^{-i}, T_i)) \geq \varepsilon \tanh(\varepsilon/2)/(2 \ln 2) \tag{4.33}$$

On the other hand, applying Theorem 4.3, there exists $\underline{t}^{-i}$ such that the constant strategy of always outputting 0 satisfies:

$$\mathbb{E}[v(T_i, \hat{M}(\underline{t}^{-i}, 0))] \geq \mathbb{E}[v(T_i, \hat{M}(\underline{t}^{-i}, T_i))] - O(\cosh(\varepsilon/2)/n^{1/4}) - \eta \tag{4.34}$$

Of course, the constant strategy of always outputting 0 reveals no information. Therefore, letting σ_0 denote the strategy that always outputs 0 and letting σ_I denote the truthful strategy, we have that:

$$u(T_i, \underline{t}^{-i}, \sigma_I) \leq u(T_i, \underline{t}^{-i}, \sigma_0) - \nu_i \varepsilon \tanh(\varepsilon/2)/(2 \ln 2) + O(\cosh(\varepsilon/2)/n^{1/4}) + \eta$$

By our hypothesis that $\nu_i = \omega(1/n^{1/4})$, this implies that for sufficient large n, player would i prefer the constant 0 strategy to the truthful strategy. ∎

Recall that typically we think of η being negligible in n, so $\eta = o(n^{-1/4})$ is satisfied. Observe that the entropy $H(T_i) = O(\log q) = O(\log n)$. By Remark 4.1, a reasonable setting of ν_i is $\nu_i = \Theta(1/\log n)$, which is much larger than $1/n^{1/4}$ and therefore satisfies the hypothesis of Theorem 4.7. Therefore, Theorem 4.7 says that for this game and for realistic parameters, revealing the entire perturbed histogram would sacrifice truthfulness.

5 Acknowledgments

Part of this work was done while the author was visiting Microsoft Research Asia. The author would like to thank Pinyan Lu and Yajun Wang for their collaboration at the initial stages of this work. The author was supported by the French ANR Defis program under contract ANR-08-EMER-012.

References

[1] Noga Alon, Michal Feldman, Ariel D. Procaccia, and Moshe Tennenholtz. Strategyproof approximation of the minimax on networks. *Mathematics of Operations Research*, 35(3):513–526, 2010.

[2] Boaz Barak, Kamalika Chaudhuri, Cynthia Dwork, Satyen Kale, Frank Mcsherry, and Kunal Talwar. Privacy, accuracy, and consistency too: a holistic solution to contingency table release. In *Proc. of 26th PODS*, pages 273–282, 2007.

[3] Avrim Blum, Katrina Ligett, and Aaron Roth. A learning theory approach to non-interactive database privacy. In *Proc. 40'th STOC*, pages 609–618, 2008.

[4] T. Dalenius. Towards a methodology for statistical disclosure control. *Statistik Tidskrift*, 5:429–444, 1977.

[5] Irit Dinur and Kobbi Nissim. Revealing information while preserving privacy. In *In PODS*, pages 202–210. ACM Press, 2003.

[6] Cynthia Dwork. Differential privacy. In *In Proc. ICALP*, pages 1–12. Springer, 2006.

[7] Cynthia Dwork. Differential privacy: A survey of results. In Manindra Agrawal, Dingzhu Du, Zhenhua Duan, and Angsheng Li, editors, *Theory and Applications of Models of Computation*, volume 4978 of *Lecture Notes in Computer Science*, pages 1–19. Springer Berlin / Heidelberg, 2008.

[8] Cynthia Dwork, Krishnaram Kenthapadi, Frank Mcsherry, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *In EUROCRYPT*, pages 486–503. Springer, 2006.

[9] Cynthia Dwork, Frank Mcsherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *In Proc. of the 3rd TCC*, pages 265–284. Springer, 2006.

[10] Cynthia Dwork, Moni Naor, Omer Reingold, Guy N. Rothblum, and Salil Vadhan. On the complexity of differentially private data release: Efficient algorithms and hardness results. In *Proc. 41'st STOC*, STOC '09, pages 381–390. ACM, 2009.

[11] Joan Feigenbaum, Aaron D. Jaggard, and Michael Schapira. Approximate privacy: foundations and quantification (extended abstract). In *Proc. 11th EC*, EC '10, pages 167–178, New York, NY, USA, 2010. ACM.

[12] Dan Feldman, Amos Fiat, Haim Kaplan, and Kobbi Nissim. Private coresets. In *Proc. 41st STOC*, STOC '09, pages 361–370, New York, NY, USA, 2009. ACM.

[13] I. Fellegi. On the question of statistical confidentiality. *J. of the Amer. Stat. Assoc.*, 67:7–18, 1972.

[14] Arpita Ghosh and Aaron Roth. Selling privacy at auction, 2010. Manuscript.

[15] Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. In *Proc. 41'st STOC*, STOC '09, pages 351–360. ACM, 2009.

[16] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? In *Proc. 49th FOCS*, pages 531–540, Washington, DC, USA, 2008. IEEE.

[17] Pinyan Lu, Xiaorui Sun, Yajun Wang, and Zeyuan Allen Zhu. Asymptotically optimal strategy-proof mechanisms for two-facility games. In *Proceedings of the 11th ACM conference on Electronic commerce*, EC '10, pages 315–324. ACM, 2010.

[18] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *Proceedings of the 48th Annual Symposium on Foundations of Computer Science*. Citeseer, 2007.

[19] H. Moulin. On strategy-proofness and single peakedness. *Public Choice*, 35:437–455, 1980.

[20] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *Proc. 39th STOC*, pages 75–84, 2007.

[21] Kobbi Nissim, Rann Smorodinsky, and Moshe Tennenholtz. Approximately Optimal Mechanism Design via Differential Privacy. *Arxiv preprint arXiv:1004.2888*, 2010.

[22] M. R. Spiegel. *Theory and Problems of Probability and Statistics*. McGraw-Hill, 1992.

A Omitted proofs

A.1 Proof of Lemma 2.5

Lemma 2.5 (Restated). *For Y, Y' independently chosen from $\mathcal{NB}_{q,\varepsilon}$, it holds that*

$$\Pr[Y = Y'] \leq \frac{4\sinh(\varepsilon/2)}{\sqrt{q-1}}$$

Proof. $\Pr[Y = Y']$ is just the collision probability of $\mathcal{NB}_{q,\varepsilon}$. By definition, the second Renyi entropy $H_2(Y) = -\log \Pr[Y = Y']$, and it is well-known that $H_2(Y) \geq H_\infty(Y) = -\log \max_{y\in\mathbb{Z}} \Pr[Y = y]$. Therefore it suffices to upper bound $\max_{y\in\mathbb{Z}} \Pr[Y = y]$.

Let $\alpha = e^{-\varepsilon}$. It is a standard fact about the negative binomial that $\max_{y\in\mathbb{Z}} \Pr[Y = y]$ is achieved at $y_0 = (q-1)\frac{\alpha}{1-\alpha}$ (assuming for simplicity that this is an integer) [22]. Therefore we have:

$$\begin{aligned}\Pr[Y = y_0] &= \binom{(q-1)\frac{\alpha}{1-\alpha} + (q-1)}{q-1}(1-\alpha)^q\alpha^{(q-1)\alpha/(1-\alpha)} \\ &= \binom{\left(\frac{q-1}{1-\alpha}\right)}{q-1}(1-\alpha)^q\alpha^{(q-1)\alpha/(1-\alpha)}\end{aligned}$$

Applying Stirling's approximation (making some gross approximations to make the notation simpler), the previous is bounded by:

$$\leq 2\frac{\left(\frac{q-1}{1-\alpha}\right)^{(q-1)/(1-\alpha)+1/2}}{(q-1)^{(q-1)+1/2}\left(\frac{(q-1)\alpha}{1-\alpha}\right)^{(q-1)\alpha/(1-\alpha)+1/2}}(1-\alpha)^q\alpha^{(q-1)\alpha/(1-\alpha)}$$

$$= 2\frac{(q-1)^{(q-1)/(1-\alpha)+1/2}\cdot\left(\frac{1}{1-\alpha}\right)^{(q-1)/(1-\alpha)+1/2}}{(q-1)^{(q-1)/(1-\alpha)+1}\left(\frac{\alpha}{1-\alpha}\right)^{(q-1)\alpha/(1-\alpha)+1/2}}(1-\alpha)^q\alpha^{(q-1)\alpha/(1-\alpha)}$$

$$= 2\frac{\left(\frac{1}{1-\alpha}\right)^{(q-1)}}{(q-1)^{1/2}\alpha^{(q-1)\alpha/(1-\alpha)+1/2}}(1-\alpha)^q\alpha^{(q-1)\alpha/(1-\alpha)}$$

$$= 2\frac{1-\alpha}{(q-1)^{1/2}\alpha^{1/2}}$$

$$= 4\frac{e^{\varepsilon/2} - e^{-\varepsilon/2}}{2\sqrt{q-1}}$$

$$= 4\frac{\sinh(\varepsilon/2)}{\sqrt{q-1}}$$

■

B Mechanisms need only consider the histogram

Suppose M is an arbitrary mechanism, possibly looking at individual types. We transform it into a mechanism M' that considers only the histogram according to Algorithm B.1.

M' is efficient because for all π it holds that the histogram $\pi(\underline{h})$ is exactly $\underline{h}$. Since the outcome of M is efficient on $\pi(\underline{h})$, therefore the outcome is efficient for $\underline{h}$.

To see that M' is truthful, suppose not and that there is an input $\underline{h}$ and a deviation that allows one player to improve his utility on this input. By an averaging argument, this means that there exists π such that there is a deviation that allows one player to improve his utility for the input $\pi(\underline{h})$ and with the mechanism M. But this contradicts the truthfulness of M.

Input: histogram $\underline{h}$ with m coordinates. Let $n = \sum_{i=1}^{m} h_i$.

1. Select a random permutation $\pi : [n] \to [n]$.
2. For $i = 1$ to n, set $t_{\sigma(i)} = \min_{j \in [m]} \sum_{k=1}^{j} \geq i$. By abuse of notation, let $\pi(h)$ denote this setting of $t_1, \ldots, t_n$.
3. Run $M(t_1, \ldots, t_n)$.

Algorithm B.1. Converting an arbitrary mechanism to one looking only at histogram.