Mechanism Design in Large Games: Incentives and Privacy *

[Extended Abstract]†

Michael Kearns
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA
mkearns@cis.upenn.edu

Mallesh M. Pai
Department of Economics
University of Pennsylvania
Philadelphia, PA
mallesh@econ.upenn.edu

Aaron Roth
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA
aaroth@cis.upenn.edu

Jonathan Ullman
School of Engineering and Applied Sciences
Harvard University
Cambridge, MA
jullman@seas.harvard.edu

ABSTRACT

We study the problem of implementing equilibria of complete information games in settings of incomplete information, and address this problem using "recommender mechanisms." A recommender mechanism is one that does not have the power to enforce outcomes or to force participation, rather it only has the power to suggestion outcomes on the basis of voluntary participation. We show that despite these restrictions, recommender mechanisms can implement equilibria of complete information games in settings of incomplete information under the condition that the game is *large*—i.e. that there are a large number of players, and any player's action affects any other's payoff by at most a small amount.

Our result follows from a novel application of *differential privacy*. We show that any algorithm that computes a correlated equilibrium of a complete information game while satisfying a variant of differential privacy—which we call *joint differential privacy*—can be used as a recommender mechanism while satisfying our desired incentive properties. Our main technical result is an algorithm for computing a correlated equilibrium of a large game while satisfying joint differential privacy.

Although our recommender mechanisms are designed to satisfy game-theoretic properties, our solution ends up satisfying a strong privacy property as well. No group of players can learn "much" about the type of any player outside the group from the recommendations of the mechanism, even if these players collude in an arbitrary way. As such, our algorithm is able to implement equilibria of complete information games, without revealing information about the realized types.

Categories and Subject Descriptors

Theory of Computation [**Algorithmic Game Theory and Mechanism Design**]: Algorithmic Mechanism Design

Keywords

Game Theory, Mechanism Design, Large Games, Differential Privacy

1. INTRODUCTION

A useful simplification common in game theory is the model of *games of complete (or full) information*. Informally, in a game of complete information, each player knows with certainty the utility function of every other player. In games of complete information, there are a number of solution concepts at our disposal, such as Nash equilibrium and correlated equilibrium. Common to these is the idea that each player is playing a best response against his opponents—because of randomness, players might be uncertain about what actions their opponents are taking, but they understand their opponents' incentives exactly.

In many situations, it is unreasonable to assume that players have exact knowledge of each other's utilities. For example, players may have few means of communication outside of the game, or may regard their type as valuable private information. These are *games of incomplete (or partial) in-*

*We gratefully acknowledge the support of NSF Grant CCF-1101389. We thank Nabil Al-Najjar, Eduardo Azevdeo, Eric Budish, Tymofiy Mylovanov, Andy Postlewaite, Al Roth and Tim Roughgarden for helpful comments and discussions.

†A full version of this working paper appears on the arXiv preprint site [KPRU13].

ITCS'14, January 12–14, 2014, Princeton, New Jersey, USA.

http://dx.doi.org/10.1145/2554797.2554834.

formation, which are commonly modeled as Bayesian games in which the players' utilities functions, or types, are drawn from a commonly known prior distribution. The most common solution concept in such games is Bayes-Nash Equilibrium. This stipulates that every player i, as a function of his type, plays an action that maximizes his expected payoff, in expectation *both* over the random draw of the types of his opponents from the prior distribution and over the possible randomization of the other players.

Unsurprisingly, equilibrium welfare can suffer in games of incomplete information, because coordination becomes more difficult amongst players who do not know each other's types. One way to measure the quality of equilibria is via the "price of anarchy"—how much worse the social welfare can be in an equilibrium outcome, as opposed to the welfare-maximizing outcome. The price of anarchy can depend significantly on the notion of equilibrium. For example, Roughgarden [Rou12] notes that even smooth games[1] that have a constant price of anarchy under full information solution concepts (e.g. Nash or correlated equilibrium) can have an unboundedly large price of anarchy under partial information solution concepts (e.g. Bayes-Nash Equilibrium). Therefore, given a game of partial information, where all that we can predict is that players will choose some Bayes-Nash Equilibrium (if even that), it may be preferable to implement an equilibrium of the complete information game defined by the actual realized types of the players. Doing so would guarantee welfare bounded by the price of anarchy of the full information game, rather than suffering from the large price of anarchy of the partial information setting. In a smooth game, we would be just as happy implementing a correlated equilibrium as a Nash equilibrium, since the price of anarchy is no worse over correlated equilibria.

In this paper we ask whether it is possible to help coordinate on an equilibrium of the realized full information game using a certain type of proxy that we call a "recommender mechanism." That is, we augment the game with an additional option for each player to use a proxy. If players opt in to using the proxy, and reveal their type to the proxy, then it will suggest some action for them to take. However, players may also simply opt out of using the proxy and play the original game using any strategy they choose. We make the assumption that if players use the proxy, then they must report their type truthfully (or, alternatively, that the proxy has the ability to verify a player's type and punish those who report dishonestly). However, the proxy has very limited power in other respects, because it does not have the ability to modify payoffs of the game (i.e. make payments) or to enforce that its recommendations be followed.

Our main result is that it is indeed possible to implement approximate correlated equilibria of the realized full information game using recommender mechanisms, assuming the original game is "large". Informally, a game is large if there are many players and that each player has individually only a small affect on the utility of any other player. We show that in such games there exists a recommender mechanism such that for any prior on agent types, it is an approximate Bayes-Nash equilibrium for every agent in the game to opt in to the proxy, and then follow its recommended action. Moreover, when players do so, the resulting play forms an approximate correlated equilibrium of the full information game. The approximation error we require tends to 0 as the number of players grows.

[1]Of particular interest to us are *smooth games*, defined by Roughgarden [Rou09]. Almost all known price of anarchy bounds (including those for the well studied model of traffic routing games) are bounds on smooth games, and many are quite good. Although price of anarchy bounds are typically proven for exact Nash equilibria of the full information games, in smooth games, the price of anarchy bounds extend without loss to (and even beyond) approximate correlated equilibria, again of the full information game.

1.1 Overview of Techniques and Results

A tempting approach is to use the following form of proxy: The proxy accepts a report of each agent's type, which defines an instance of a full information game. It then computes a correlated equilibrium of the full information game, and suggests an action to each player which is a draw from this correlated equilibrium. By definition of a correlated equilibrium, if all players opt into the proxy, then they can do no better than subsequently following the recommended action. However, this proxy does not solve the problem, as it may not be in a player's best interest to opt in, even if the other $n-1$ players do opt in! Intuitively, by opting out, the player can cause the proxy to compute a correlated equilibrium of the wrong game, or to compute a different correlated equilibrium of the same game![2] The problem is an instance of the well known equilibrium-selection problem—even in a game of full information, different players may disagree on their preferred equilibrium, and may have trouble coordinating. The problem is only more difficult in settings of incomplete information. In our case, by opting out of the mechanism, a player can have a substantial affect on the computed equilibrium, even if each player has only small affect on the utilities of other players.

Our solution is to devise a means of computing correlated equilibria such that any single player's reported type to the algorithm only has a small effect on the distribution of suggested actions to all other players. The precise notion of "small effect" that we use is a variant of the well studied notion of *differential privacy*. It is not hard to see that computing an equilibrium of even a large game is not

[2]As a simple example, consider a large number n of people who must each choose whether to go to the beach (B) or mountains (M). People privately know their types— each person's utility depends on his own type, his action, and the fraction of other people p who go to the beach. A Beach type gets a payoff of $10p$ if he visits the beach, and $5(1-p)$ if he visits the mountain. A mountain type gets a payoff $5p$ from visiting the beach, and $10(1-p)$ from visiting the mountain. Note that the game is 'insensitive' (an agent's visit decision has a small impact on others' payoffs). Further, note that "everyone visits beach" and "everyone visits mountain" are both equilibria of the game, regardless of the realization of types. Consider the mechanism that attempts to implement the following social choice rule—"if the number of beach types is less than half the population, send everyone to the beach, and vice versa." It should be clear that if mountain types are just in the majority, then each mountain type has an incentive to opt out of the mechanism, and vice versa. As a result, even though the game is "large" and agents' actions do not affect others' payoffs significantly, simply computing equilibria from reported type profiles does not in general lead to even approximately truthful mechanisms. This is a general phenomenon that is not specific to our example. Finding an exact correlated equilibrium subject to any objective is a linear programming problem, and in general small changes in the objective (or constraints) of an LP can lead to wild changes in its solution.

possible under the standard constraint of differential privacy, because although agent's actions have only a small affect on the utilities of other players in large games, they can have large affect on their own utility functions. Thus, it is not possible to privately announce a best response for player i while protecting the privacy of i's type. Instead, we introduce a variant which we call *joint differential privacy*, which requires that simultaneously for every player i, the joint distribution on the suggested actions to all players $j \neq i$ be differentially private in the type of agent i. We show that a proxy mechanism which calculates an α-approximate correlated equilibrium of the game induced by players reported types, under the constraint of ϵ-joint differential privacy makes it an $(\epsilon + \alpha)$-approximate Bayes-Nash equilibrium for players to opt into the proxy, and then follow their suggested action, as desired.

Our main technical result is an instantiation of this plan: a pair of algorithms for computing α-approximate correlated equilibria in large games, such that we can take the approximation parameter $\epsilon + \alpha$ tending to zero. The first algorithm is efficient, but has a suboptimal dependence on the number of actions k in the game. The other algorithm is inefficient, but has a nearly optimal dependence on k. Both have an optimal dependence on the number of players n in the game, which we show by exhibiting a matching lower bound.

We introduce joint differential privacy, large games, and our game theoretic solution concepts in Section 2. In Section 3, we formally introduce our notion of a proxy, and state our main results. Formal proofs are deferred to the full working paper version available at on the arXiv preprint site [KPRU13].

1.2 Related Work and Discussion

Market and Mechanism Design Our work is related to the large body of literature on mechanism/ market design in "large games," which uses the large number of agents to provide mechanisms which have good incentive properties, even when the small market versions do not. It stretches back to [RP76] who showed that market (Walrasian) equilibria are approximately strategy proof in large economies. More recently [IM05], [KP09], [KPR10] have shown that various two-sided matching mechanisms are approximately strategy proof in large markets. There are similar results in the literature for one-sided matching markets, market economies, and double auctions. The most general result is that of [AB11] who design incentive compatible mechanisms for large economies that satisfy a smoothness assumption. While we only allow agents to opt in/ opt out rather than mis-report, we do not assume any such smoothness condition. Further, the literature on mechanism design normally gives the mechanism the power to "enforce" actions, while here our mechanism can only "recommend" actions.

Our work is also related to mediators in games [MT03, MT09]. This line of work aims to modify the equilibrium structure of full information games by introducing a mediator, which can coordinate agent actions if they choose to opt in using the mediator. Mediators can be used to convert Nash equilibria into dominant strategy equilibria [MT03], or implement equilibrium that are robust to collusion [MT09]. Our notion of a recommender mechanism is related, but is even weaker than that of a mediator. For example, our mechanisms do not need the power to make payments [MT03], or the power to enforce suggested actions [MT09]. Our mediators are thus closer to the communication devices in the "communication equilibria" of Forges [For86]—that work investigates the set of achievable payoffs via such mediators rather than how to design one, which we do here. It also does not allow players to opt out of using the mediator.

Large Games Our results hold under a "largeness condition", i.e. a player's action affects the payoff of all others by a small amount. These are closely related to the literature on large games, see e.g. [ANS00] or [Kal04]. There has been recent work studying large games using tools from theoretical computer science (but in this case, studying robustness of equilibrium concepts)—see [GR08, GR10].

Differential Privacy Differential privacy was first defined by [DMNS06], and is now the standard privacy "solution concept" in the theoretical computer science literature. It quantifies the *worst-case harm* that can befall an individual from allowing his data to be used in a computation, as compared to if he did not provide his data. There is by now a very large literature on differential privacy, readers can consult [Dwo08] for a more thorough introduction to the field.

[MT07] were the first to observe that a differentially private algorithm is also approximately truthful. This line of work was extended by [NST12] to give mechanisms in several special cases which are exactly truthful by combining private mechanisms with non-private mechanisms which explicitly punish non-truthful reporting. [HK12] showed that the mechanism of [MT07] (the "exponential mechanism") is in fact maximal in distributional range, and so can be made exactly truthful with the addition of payments. This immediate connection between privacy and truthfulness does not carry over to the notion of joint-differential privacy that we study here, but as we show, it is regained if the object that we compute privately is an equilibrium of the underlying game.

Another interesting line of work considers the problem of designing truthful mechanisms for agents who explicitly experience a cost for privacy loss as part of their utility function [CCK+13, NOS12, Xia13]. The main challenge in this line of work is to formulate a reasonable model for how agents experience cost as a function of privacy. We remark that the approaches taken in the former two can also be adapted to work in our setting, for agents who explicitly value privacy. [Gra12] studies the problem of implementation for various assumptions about players' preference for privacy and permissible game forms. A related line of work which also takes into account agent values for privacy considers the problem of designing markets by which analysts can procure private data from agents who explicitly experience costs for privacy loss [FL12, GR11, LR12, RS12]. See [PR13] for a survey.

Finally, our results rely on several technical solutions in the differential privacy literature. The most well studied problem is that of accurately answering numeric-valued queries on a data set. A basic result of [DMNS06] is that any low sensitivity query (i.e. the addition or removal of a single entry can change the value of the query by at most 1) can be answered efficiently and (ϵ-differential) privately while introducing only $O(1/\epsilon)$ error. Another fundamental result of [DKM+06, DRV10] is that differential privacy composes gracefully. Any algorithm composed of T subroutines, each of which are $O(\epsilon)$-differentially private, is itself $\sqrt{T}\epsilon$-

differentially private. Combined, these give an efficient algorithm for privately answering any T low sensitivity queries with $O(\sqrt{T})$ effort, a result which we make use of.

Using computationally inefficient algorithms, it is possible to privately answer queries much more accurately [BLR08, DRV10, RR10, HR10, GHRU11, GRU12]. Combining the results of the latter two yields an algorithm which can privately answer arbitrary low sensitivity queries as they arrive, with error that scales only logarithmically in the number of queries. We use this when we consider games with large action spaces.

Our lower bounds for privately computing equilibria use recent information theoretic lower bounds on the accuracy queries can be answered while preserving differential privacy [DN03, DMT07, DY08, De12]. Namely, we construct games whose equilibria encode answers to large numbers of queries on a database.

Variants of differential privacy related to joint differential privacy have been considered in the setting of query release, specifically for analyst privacy [DNV12]. Specifically, the definition of one-analyst-to-many-analyst privacy used by [HRU13] can be seen as an instantiation of joint differential privacy.

2. MODEL & PRELIMINARIES

We consider games $\mathcal{G}$ of up to n players $\{1, 2, \ldots, n\}$, indexed by i. Player i can take actions in a set A, $|A| = k$. To allow our games to be defined also for fewer than n players, we will imagine that the null action $\perp \in A$, which corresponds to "opting out" of the game. We index actions by j. A tuple of actions, one for each player, will be denoted $\mathbf{a} = (a_1, a_2, \ldots a_n) \in A^n$.[3]

Let $\mathcal{U}$ be the set of player types.[4] There is a utility function $u : \mathcal{U} \times A^n \to \Re$ that determines the payoff for a player given his type t_i and a joint action profile $\mathbf{a}$ for all players. When it is clear from context, we will refer to the utility function of player i, writing $u_i : A^n \to \Re$ to denote $u(t_i, \cdot)$. We write a generic profile of utilities $\mathbf{u} = (u_1, u_2, \ldots u_n) \in \mathcal{U}^n$. We will be interested in implementing equilibria of the *complete information game* in settings of *incomplete information*. In the complete information setting, the types t_i of each player is fixed and commonly known to all players. In such settings, we can ignore the abstraction of 'types' and consider each player i simply to have a fixed utility function u_i. In models of *incomplete* information, players know their own type, but do not know the types of others. In the Bayesian model of incomplete information, there is a commonly known prior distribution τ from which each agent's type is jointly drawn: $(t_1, \ldots, t_n) \sim \tau$. We now define the solution concepts we will use, both in the full information setting and in the Bayesian setting.

Denote a distribution over A^n by π, the marginal distribution over the actions of player i by π_i, and the marginal distribution over the (joint tuple of) actions of every player but player i by π_{-i}. We now present two standard solution concepts— approximate correlated and coarse correlated equilibrium.

[3]In general, subscripts will refer indices i.e. players and periods, while superscripts will refer to components of vectors.

[4]It is trivial to extend our results when agents have different typesets, $\mathcal{U}_i$. $\mathcal{U}$ will then be $\bigcup_{i=1}^{n} \mathcal{U}_i$.

DEFINITION 1 (APPROXIMATE CE). *Let $(u_1, u_2, \ldots u_n)$ be a tuple of utility functions, one for each player. Let π be a distribution over tuples of actions A^n. We say that π is an* α-approximate correlated equilibrium *of the (complete information) game defined by $(u_1, u_2, \ldots u_n)$ if for every player $i \in [N]$, and any function $f : A \to A$,*

$$\mathop{\mathbb{E}}_{\pi}[u_i(\mathbf{a})] \geq \mathop{\mathbb{E}}_{\pi}[u_i(f(a_i), a_{-i})] - \alpha$$

We now define a solution concept in the Bayesian model. Let τ be a commonly known joint distribution over $\mathcal{U}^n$, and let $\tau_{|t_i}$ be the posterior distribution on types conditioned on the type of player i being t_i. A (pure) *strategy* for player i is a function $s_i : \mathcal{U} \to A$, and we write $\mathbf{s} = (s_1, \ldots, s_n)$ to denote a vector of strategy profiles.

DEFINITION 2 (APPROXIMATE BNE). *Let τ be a distribution over $\mathcal{U}^n$, and let $\mathbf{s} = (s_1, \ldots, s_n)$ be a vector of strategies. We say that $\mathbf{s}$ is an* α-approximate (pure strategy) Bayes Nash Equilibrium *under τ if for every player i, for every $t_i \in \mathcal{U}$, and for every alternative strategy s_i':*

$$\mathop{\mathbb{E}}_{\mathbf{t}_{-i} \sim \tau_{|t_i}}[u_i(t_i, s_i(t_i), s_{-i}(t_{-i}))] \geq$$

$$\mathop{\mathbb{E}}_{\mathbf{t}_{-i} \sim \tau_{|t_i}}\left[u_i(t_i, s_i'(t_i), s_{-i}(t_{-i}))\right] - \epsilon$$

We restrict attention to 'insensitive' games. Roughly speaking a game is γ-sensitive if a player's choice of action affects any other player's payoff by at most γ. Note that we do not constrain the effect of a player's *own* actions on his payoff— *a player's action can have a large impact on his own payoff.* Formally:

DEFINITION 3 (γ-SENSITIVE). *A game is said to be γ-sensitive if for any two distinct players $i \neq i'$, any two actions a_i, a_i' and type t_i for player i and any tuple of actions a_{-i} for everyone else:*

$$|u_{i'}(a_i, a_{-i}) - u_{i'}(a_i', a_{-i})| \leq \gamma. \quad (1)$$

A key tool in our paper is the design of differentially private "proxy" algorithms for suggesting actions to play. Agents can able to opt out of participating in the proxy: so each agent can submit to the proxy either their type t_i, or else a null symbol $\perp$ which represents opting out. A proxy algorithm is then a function from a profile of utility functions (and $\perp$ symbols) to a probability distribution over $\mathcal{R}^n$, i.e. $\mathcal{M} : (\mathcal{U} \cup \{\perp\})^n \to \Delta\mathcal{R}^n$. Here $\mathcal{R}$ is an appropriately defined range space.

First we recall the definition of differential privacy, both to provide a basis for our modified definition, and since it will be a technical building block in our algorithms. Roughly speaking, a mechanism is differentially private if for every $\mathbf{u}$ and every i, knowledge of the output $\mathcal{M}(\mathbf{u})$ as well as u_{-i} does not reveal 'much' about u_i.

DEFINITION 4 ((STANDARD) DIFFERENTIAL PRIVACY). *A mechanism $\mathcal{M}$ satisfies (ε, δ)-differential privacy if for any player i, any two types for player i, t_i and $t_i' \in \mathcal{U} \cup \{\perp\}$, and any tuple of types for every else $t_{-i} \in (\mathcal{U} \cup \{\perp\})^{n-1}$ and any $S \subseteq \mathcal{R}^n$,*

$$\mathop{\mathbb{P}}_{\mathcal{M}}[(\mathcal{M}(t_i; t_{-i})) \in S] \leq e^{\varepsilon} \mathop{\mathbb{P}}_{\mathcal{M}}\left[\left(\mathcal{M}(t_i'; t_{-i})\right) \in S\right] + \delta.$$

We would like something slightly different for our setting. We propose a relaxation of the above definition, motivated by the fact that when computing a correlated equilibrium, the action recommended to a player is only observed by her. Roughly speaking, a mechanism is *jointly differentially private* if, for each player i, knowledge of the other $n-1$ recommendations (and submitted types) does not reveal 'much' about player i's report. Note that this relaxation is necessary in our setting if we are going to privately compute correlated equilibria, since knowledge of player i's recommended action can reveal a lot of information about his type. It is still very strong- the privacy guarantee remains *even if* everyone else colludes against a given player i, so long as i does not himself make the component reported to him public.

DEFINITION 5 (JOINT DIFFERENTIAL PRIVACY). *A mechanism $\mathcal{M}$ satisfies (ε, δ)-joint differential privacy if for any player i, any two possible types for player i, t_i and $t_i' \in \mathcal{U} \cup \{\perp\}$, any tuple of utilities for everyone else t_{-i} and $S \subseteq \mathcal{R}^{n-1}$,*

$$\mathop{\mathbb{P}}_{\mathcal{M}}\left[(\mathcal{M}(t_i; t_{-i}))_{-i} \in S\right] \leq e^{\varepsilon} \mathop{\mathbb{P}}_{\mathcal{M}}\left[\left(\mathcal{M}(t_i'; t_{-i})\right)_{-i} \in S\right] + \delta.$$

3. JOINT DIFFERENTIAL PRIVACY AND TRUTHFULNESS

The main result of this paper is a reduction that takes an arbitrary large game $\mathcal{G}$ of incomplete information and modifies it to have equilibrium implementing equilibrium outcomes of the corresponding full information game defined by the *realized* agent types. Specifically, we modify the game by introducing the option for players to use a *proxy* that can recommend actions to the players. The modified game is called $\mathcal{G}'$. For any prior on agent types, it will be an approximate Bayes Nash equilibrium of $\mathcal{G}'$ for every player to opt in to using the proxy, and to subsequently follow its recommendation. Moreover, the resulting set of actions will correspond to an approximate correlated equilibrium of the complete information game $\mathcal{G}$ defined by the realized agent types. For concreteness, we consider Bayesian games, however our results are not specific to this model of incomplete information.

More precisely, the modified game $\mathcal{G}'$ will be identical to $\mathcal{G}$ with an added option. Each player i has the opportunity to submit their type to a proxy, which will then suggest to them an action $\hat{a}_i \in A$ to play. They can use this advice however they like: that is, they can choose any function $f : A \to A$ and choose to play the action $a_i = f(\hat{a}_i)$. Alternately, they can opt out of the proxy (and not submit their type), and choose an action to play $a_i \in A$ directly. In the end, each player experiences utility $u(t_i, (a_1, \ldots, a_n))$, just as in the original game $\mathcal{G}$. We assume that types are verifiable—agent i does not have the ability to opt in to the proxy but report a false type $t_i' \neq t_i$. However, he does have the ability to opt out (and submit $\perp$), and the proxy has no power to do anything other than suggest which action he should play. In the end, each player is free to play any action a_i, regardless of what the proxy suggests, even if he opts in.

Formally, given a game $\mathcal{G}$ defined by an action set A, a type space $\mathcal{U}$, and a utility function u, we define a *proxy game* $\mathcal{G}'_M$, parameterized by an algorithm $M : \{\mathcal{U} \cup \{\perp\}\}^n \to A^n$. In $\mathcal{G}'$, each agent has two types of actions: they can *opt in* to the proxy, which means they submit their type, receive an action recommendation $\hat{a}$, and choose a function $f : A \to A$ which determines how they use that recommendation. We denote this set of choices $A_1' = \{(\top, f) | f : A \to A\}$. Alternately, they can *opt out* of the proxy, which means that they do not submit their type, and directly choose an action to play. We denote this set of choices $A_2' = \{(\perp, a) | a \in A\}$. Together, the action set in $\mathcal{G}'_M$ is $A' = A_1' \cup A_2'$. Given a set of choices by the players, we define a vector $\mathbf{x}$ such that $\mathbf{x}_i = t_i$ for each player i who chose $(\top, f_i) \in A_1'$ (each player who opted in), and $\mathbf{x}_i = \perp$ for each player i who chose $(\perp, a_i) \in A_2'$ (each player who opted out). The proxy then computes $M(\mathbf{x}) = \hat{\mathbf{a}}$. Finally, this results in a vector of actions $\mathbf{a}$ from the game $\mathcal{G}$, one for each player. For each player who opted in, they play the action $\mathbf{a}_i = f_i(\hat{\mathbf{a}}_i)$. For each person who opted out, they play the action $\mathbf{a}_i = a_i$. Finally, each player receives utility $u(t_i, \mathbf{a})$ as in the original game $\mathcal{G}$.

We now show that if the algorithm M satisfies certain properties, then for any prior on agent types, it is always an approximate Bayes Nash equilibrium for every player to opt in and follow the recommendation of the proxy.

THEOREM 6. *Let M be an algorithm that satisfies (ϵ, δ)-joint differential privacy, and be such that for every vector of types $\mathbf{t} \in \mathcal{U}^n$, $M(\mathbf{t})$ induces a distribution over actions that is an α-approximate correlated equilibrium of the full information game $\mathcal{G}$ induced by the type vector $\mathbf{t}$. Then for every prior distribution on types τ, it is an η-approximate Bayes Nash equilibrium of $\mathcal{G}'_M$ for every player to play $(\top, f)$ for the identity function $f(a) = a$. (i.e. for every player to opt into the proxy, and then follow its suggested action), where $\eta = \epsilon + \delta + \alpha$.*

REMARK 7. *Observe that when agents play according to the approximate Bayes Nash equilibrium of $\mathcal{G}'_M$ guaranteed by Theorem 6, then the resulting distribution over actions played, and the resulting utilities of the players, correspond to an α-approximate correlated equilibrium of the full information game $\mathcal{G}'_M$, induced by the realized type vector $\mathbf{t}$.*

PROOF PROOF OF THEOREM 6. Fix any prior distribution on player types τ, and let $s_1, \ldots, s_n$ be the strategies corresponding to the action $(\top, f)$ for each player, where f is the identify function. (i.e. the strategy corresponding to opting into the proxy and following the suggested action). There are two types of deviations that a player i can consider: $(\top, f_i')$ for some function $f_i' : A \to A$ not the identity function, and $(\perp, a_i)$ for some action a_i. First, we consider deviations of the first kind. Let $s_i'(t_i)$ be the strategy corresponding to playing $(\top, f'_{\hat{s}(t_i)})$ for some function $\hat{s}(t_i)$. For every type t_i:

$$\mathop{\mathbb{E}}_{\mathbf{t}_{-i} \sim \tau_{|t_i}}\left[u_i(t_i, s_i(t_i), s_{-i}(t_{-i}))\right] = \sum_{\mathbf{t}_{-i}} \Pr_{\tau_{|t_i}}[\mathbf{t}_{-i}] \cdot \mathop{\mathbb{E}}_{\mathbf{a} \sim M(\mathbf{t})}\left[u_i(\mathbf{a})\right]$$

$$\geq \sum_{\mathbf{t}_{-i}} \Pr_{\tau_{|t_i}}[\mathbf{t}_{-i}] \cdot \mathop{\mathbb{E}}_{\mathbf{a} \sim M(\mathbf{t})}\left[u_i(f'_{\hat{s}(t_i)}(\mathbf{a}_i), \mathbf{a}_{-i})\right] - \alpha$$

$$= \mathop{\mathbb{E}}_{\mathbf{t}_{-i} \sim \tau_{|t_i}}\left[u_i(t_i, s_i'(t_i), s_{-i}(t_{-i}))\right] - \alpha$$

where the inequality follows from the fact that M computes an α-approximate correlated equilibrium. Now, consider a deviation of the second kind. Let $s_i'(t_i)$ be the strategy corresponding to playing $(\perp, a_{\hat{s}(t_i)})$ for some function $\hat{s}(t_i)$. For every type t_i:

$$\mathop{\mathbb{E}}_{\mathbf{t}_{-i}\sim\tau_{|t_i}}\left[u_i(t_i, s_i(t_i), s_{-i}(t_{-i}))\right]$$

$$= \sum_{\mathbf{t}_{-i}} \Pr_{\tau_{|t_i}}[\mathbf{t}_{-i}] \cdot \mathop{\mathbb{E}}_{\mathbf{a}\sim M(\mathbf{t})}\left[u_i(\mathbf{a})\right]$$

$$\geq \sum_{\mathbf{t}_{-i}} \Pr_{\tau_{|t_i}}[\mathbf{t}_{-i}] \cdot \mathop{\mathbb{E}}_{\mathbf{a}\sim M(\mathbf{t})}\left[u_i(a_{\hat{s}(t_i)}, \mathbf{a}_{-i})\right] - \alpha$$

$$\geq \sum_{\mathbf{t}_{-i}} \Pr_{\tau_{|t_i}}[\mathbf{t}_{-i}] \cdot \exp(-\epsilon) \cdot \mathop{\mathbb{E}}_{\mathbf{a}\sim M(\perp,\mathbf{t}_{-i})}\left[u_i(a_{\hat{s}(t_i)}, \mathbf{a}_{-i})\right] - \delta - \alpha$$

$$\geq \sum_{\mathbf{t}_{-i}} \Pr_{\tau_{|t_i}}[\mathbf{t}_{-i}] \cdot \mathop{\mathbb{E}}_{\mathbf{a}\sim M(\perp,\mathbf{t}_{-i})}\left[u_i(a_{\hat{s}(t_i)}, \mathbf{a}_{-i})\right] - \epsilon - \delta - \alpha$$

$$= \mathop{\mathbb{E}}_{\mathbf{t}_{-i}\sim\tau_{|t_i}}\left[u_i(t_i, s_i'(t_i), s_{-i}(t_{-i}))\right] - \varepsilon - \delta - \alpha$$

where the first inequality follows from the α-approximate correlated equilibrium condition, the second follows from the (ϵ, δ)-joint differential privacy condition, and the third follows from the fact that for $\epsilon \geq 0$, $\exp(-\epsilon) \geq 1 - \epsilon$ and that utilities are bounded in $[0, 1]$. □

The main technical contribution of the paper is an algorithm M which satisfies (ϵ, δ)-joint differential privacy, and computes an α-approximate correlated equilibrium of games which are γ-large. We in fact give two such algorithms: one that runs in time polynomial in n and $|A| = k$, and one that runs in time exponential in n and k. The efficient algorithm computes an α_1-approximate correlated equilibrium, and the inefficient algorithm computes an α_2-approximate correlated equilibrium, where:

$$\alpha_1 = \tilde{O}\left(\frac{\gamma k^{3/2}\sqrt{n\log(1/\delta)}}{\epsilon}\right)$$

$$\alpha_2 = \tilde{O}\left(\frac{\gamma \log k \log^{3/2}(\mathcal{U})\sqrt{n}\log(1/\delta)}{\epsilon}\right).$$

In combination with Theorem 6, the existence of these algorithms together with optimal choices of ϵ and δ give our main result:

THEOREM 8. *Let $\mathcal{G}$ be any γ-large game. Then there exists a proxy game $\mathcal{G}'$ such that for any prior distribution on types τ, it is an η-approximate Bayes-Nash equilibrium to opt into the proxy and follow its advice. Moreover, the resulting distribution on actions forms an η-approximate correlated equilibrium of the full information game induced by the realized types. If we insist that the proxy be implemented using a computationally efficient algorithm, then we can take:*

$$\eta = \tilde{O}\left(\sqrt{\gamma}n^{1/4}k^{3/4}\right)$$

If we can take the proxy to be computationally inefficient, then we can take:

$$\eta = \tilde{O}\left(\sqrt{\gamma}n^{1/4}\sqrt{\log k}\log^{3/4}|\mathcal{U}|\right)$$

REMARK 9. *In large games, the parameter γ tends to zero as n grows large. For $\gamma = 1/n$, our approximation error is $\eta = \tilde{O}(k^{3/4}/n^{1/4})$ and $\eta = \tilde{O}(\sqrt{\log k}\log^{3/4}|\mathcal{U}|/n^{1/4})$ respectively. Note that the approximation error η in the equilibrium concepts tends to zero in any γ-large game such that $\gamma = o(\frac{1}{\sqrt{n}k^{3/2}})$ or $\gamma = o(\frac{1}{\sqrt{n}\log k\log^{3/2}|\mathcal{U}|})$ respectively.*

4. DISCUSSION

In this work, we have introduced a new variant of differential privacy (joint differential privacy), and have shown how it can be used as a tool to construct extremely weak proxy mechanisms which can implement equilibria of full information games, even when the game is being played in a setting of only partial information. Moreover, our privacy solution concept maintains the property that no coalition of players can learn (much) more about any player's type outside of the coalition than they could have learned in the original Bayesian game, and thus players have almost no incentive not to participate even if they view their type as sensitive information. Although our proxies are weak in most respects (they cannot enforce actions, they cannot make payments or charge fees, they cannot compel participation), we do make the assumption that player types are verifiable in the event that they choose to opt into the proxy. This assumption is reasonable in many settings: for example, in financial markets, there may be legal penalties for a firm misrepresenting relevant facts about itself, and in traffic routing games, the proxy may be embodied as a physical device (e.g. a GPS device) that can itself verify player types (e.g. physical location). Nevertheless, we view relaxing this assumption as an important direction for future work.

In this extended abstract, we have merely summarized our results. Our full paper [KPRU13] includes a formal description of our algorithms, formal statements of our theorems, and proofs of all results.

5. REFERENCES

[AB11] E. Azevedo and E. Budish. Strategyproofness in the large as a desideratum for market design. Technical report, University of Chicago, 2011.

[ANS00] N.I. Al-Najjar and R. Smorodinsky. Pivotal players and the characterization of influence. *Journal of Economic Theory*, 92(2):318–342, 2000.

[BLR08] Avrim Blum, Katrina Ligett, and Aaron Roth. A learning theory approach to non-interactive database privacy. In Cynthia Dwork, editor, *STOC*, pages 609–618. ACM, 2008.

[CCK+13] Yiling Chen, Stephen Chong, Ian A Kash, Tal Moran, and Salil Vadhan. Truthful mechanisms for agents that value privacy. In *Proceedings of the fourteenth ACM conference on Electronic commerce*, pages 215–232. ACM, 2013.

[De12] Anindya De. Lower bounds in differential privacy. In *TCC*, pages 321–338, 2012.

[DKM+06] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *EUROCRYPT*, pages 486–503, 2006.

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *TCC '06*, pages 265–284, 2006.

[DMT07] Cynthia Dwork, Frank McSherry, and Kunal Talwar. The price of privacy and the limits of lp decoding. In *STOC*, pages 85–94, 2007.

[DN03] Irit Dinur and Kobbi Nissim. Revealing information while preserving privacy. In *PODS*, pages 202–210, 2003.

[DNV12] Cynthia Dwork, Moni Naor, and Salil Vadhan. The privacy of the analyst and the power of the state. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 400–409. IEEE, 2012.

[DRV10] Cynthia Dwork, Guy N. Rothblum, and Salil P. Vadhan. Boosting and differential privacy. In *FOCS*, pages 51–60. IEEE Computer Society, 2010.

[Dwo08] C. Dwork. Differential privacy: A survey of results. *Theory and Applications of Models of Computation*, pages 1–19, 2008.

[DY08] Cynthia Dwork and Sergey Yekhanin. New efficient attacks on statistical disclosure control mechanisms. In *CRYPTO*, pages 469–480, 2008.

[FL12] Lisa Fleischer and Yu-Han Lyu. Approximately optimal auctions for selling privacy when costs are correlated with data. In *EC*, pages 568–585, 2012.

[For86] Françoise Forges. An approach to communication equilibria. *Econometrica: Journal of the Econometric Society*, pages 1375–1385, 1986.

[GHRU11] Anupam Gupta, Moritz Hardt, Aaron Roth, and Jonathan Ullman. Privately releasing conjunctions and the statistical query barrier. In *STOC '11*, pages 803–812, 2011.

[GR08] R. Gradwohl and O. Reingold. Fault tolerance in large games. In *EC*, pages 274–283. ACM, 2008.

[GR10] R. Gradwohl and O. Reingold. Partial exposure in large games. *Games and Economic Behavior*, 68(2):602–613, 2010.

[GR11] Arpita Ghosh and Aaron Roth. Selling privacy at auction. In *EC*, pages 199–208, 2011.

[Gra12] R. Gradwohl. Privacy in implementation. 2012.

[GRU12] Anupam Gupta, Aaron Roth, and Jonathan Ullman. Iterative constructions and private data release. In *TCC*, pages 339–356, 2012.

[HK12] Zhiyi Huang and Sampath Kannan. The exponential mechanism for social welfare: Private, truthful, and nearly optimal. In *FOCS*, 2012.

[HR10] Moritz Hardt and Guy N. Rothblum. A multiplicative weights mechanism for privacy-preserving data analysis. In *FOCS*, pages 61–70, 2010.

[HRU13] Justin Hsu, Aaron Roth, and Jonathan Ullman. Differential privacy for the analyst via private equilibrium computation. In *Proceedings of the 45th annual ACM symposium on Symposium on theory of computing*, pages 341–350. ACM, 2013.

[IM05] N. Immorlica and M. Mahdian. Marriage, honesty, and stability. In *SODA*, pages 53–62. Society for Industrial and Applied Mathematics, 2005.

[Kal04] E. Kalai. Large robust games. *Econometrica*, 72(6):1631–1665, 2004.

[KP09] F. Kojima and P.A. Pathak. Incentives and stability in large two-sided matching markets. *American Economic Review*, 99(3):608–627, 2009.

[KPR10] F. Kojima, P.A. Pathak, and A.E. Roth. Matching with couples: Stability and incentives in large markets. Technical report, National Bureau of Economic Research, 2010.

[KPRU13] Michael Kearns, Mallesh M Pai, Aaron Roth, and Jonathan Ullman. Mechanism design in large games: Incentives and privacy. *arXiv preprint arXiv:1207.4084*, 2013.

[LR12] Katrina Ligett and Aaron Roth. Take it or leave it: Running a survey when privacy comes at a cost. *CoRR*, abs/1202.4741, 2012.

[MT03] Dov Monderer and Moshe Tennenholtz. k-implementation. In *Proceedings of the 4th ACM conference on Electronic commerce*, pages 19–28. ACM, 2003.

[MT07] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *FOCS*, pages 94–103, 2007.

[MT09] Dov Monderer and Moshe Tennenholtz. Strong mediated equilibrium. *Artificial Intelligence*, 173(1):180–195, 2009.

[NOS12] Kobbi Nissim, Claudio Orlandi, and Rann Smorodinsky. Privacy-aware mechanism design. In *EC*, pages 774–789, 2012.

[NST12] Kobbi Nissim, Rann Smorodinsky, and Moshe Tennenholtz. Approximately optimal mechanism design via differential privacy. In *ITCS*, pages 203–213, 2012.

[PR13] Mallesh Pai and Aaron Roth. Privacy and mechanism design. *Sigecom Exchanges*, pages 8–29, 2013.

[Rou09] T. Roughgarden. Intrinsic robustness of the price of anarchy. In *STOC*, pages 513–522. ACM, 2009.

[Rou12] T. Roughgarden. The price of anarchy in games of incomplete information. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 862–879. ACM, 2012.

[RP76] D.J. Roberts and A. Postlewaite. The incentives for price-taking behavior in large exchange economies. *Econometrica*, pages 115–127, 1976.

[RR10] Aaron Roth and Tim Roughgarden. Interactive privacy via the median mechanism. In *STOC '10*, pages 765–774, 2010.

[RS12] Aaron Roth and Grant Schoenebeck. Conducting truthful surveys, cheaply. In *EC*, pages 826–843, 2012.

[Xia13] David Xiao. Is privacy compatible with truthfulness? In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 67–86. ACM, 2013.