The Exponential Mechanism for Social Welfare: Private, Truthful, and Nearly Optimal

Zhiyi Huang
Computer and Information Science
University of Pennsylvania
Philadelphia, USA
Email: hzhiyi@cis.upenn.edu

Sampath Kannan
Computer and Information Science
University of Pennsylvania
Philadelphia, USA
Email: kannan@cis.upenn.edu

***Abstract*—In this paper we show that for *any* mechanism design problem with the objective of maximizing social welfare, the exponential mechanism can be implemented as a *truthful* mechanism while still preserving differential privacy. Our instantiation of the exponential mechanism can be interpreted as a generalization of the VCG mechanism in the sense that the VCG mechanism is the extreme case when the privacy parameter goes to infinity. To our knowledge, this is the first general tool for designing mechanisms that are both truthful and differentially private.**

***Keywords*-mechanism design; differential privacy; exponential mechanism;**

I. Introduction

In mechanism design a central entity seeks to allocate resources among a set of selfish agents in order to optimize a specific objective function such as revenue or social welfare. Each agent has a private valuation for the resources being allocated, which is commonly referred to as her *type*. A major challenge in designing mechanisms for problems of resource allocation among selfish agents is getting them to reveal their true types. While in principle mechanisms can be designed to optimize some objective function even when agents are not truthful, the analysis of such mechanisms is complicated and the vast majority of mechanisms are designed to incentivize agents to be truthful.

One reason that an agent might not want to be truthful is that lying gives her a better payoff. Research in algorithmic mechanism design has mostly focused on this possibility and has successfully designed computationally-efficient *incentive-compatible* mechanisms for many problems , i.e., mechanisms where each agent achieves optimal payoff by bidding truthfully (see [23] for a survey of results). However, a second reason that an agent might not bid truthfully is that the *privacy* of her type might itself be of value to her. Bidding truthfully could well result in an outcome that reveals the private type of an agent.

The first author was supported in part by an ONR MURI Grant N000140710907. The second author was supported in part by an EAGER grant, NSF CCF 1137084.

Consider for example, a matching market in which n oil companies are bidding for n oil fields. A company may have done extensive research in figuring out its valuations for each field. It may regard this information as giving it competitive advantage and seek to protect its privacy. If it participates in a traditional incentive-compatible mechanism, say, the VCG mechanism, it has two choices – 1) bid truthfully, get the optimum payoff but potentially reveal private information or 2) introduce random noise into its bid to (almost) preserve privacy, but settle for a suboptimal payoff. In this and more generally in multi-agent settings where each agent's type is multidimensional, we aim to answer the following question:

> *Can we design mechanisms that simultaneously achieve near optimal social welfare, are incentive compatible, and protect the privacy of each agent?*

The notion of privacy we will consider is *differential privacy*, which is a paradigm for private data analysis developed in the past decade, aiming to reveal information about the population as a whole, while protecting the privacy of each individual (see surveys [13], [14] and the reference therein).

A. Our Results and Techniques

Our main contribution is a novel instantiation of the exponential mechanism for *any* mechanism design problem with payments, that aims to maximize social welfare. We show that our version of the exponential mechanism is incentive compatible and individually rational[1], while preserving differential privacy. In fact, we show that the exponential mechanism can be interpreted as a natural generalization of the VCG mechanism in the sense that the VCG mechanism is the special case when the privacy parameter goes to infinity. Alternatively, our mechanism can be viewed as an affine maximum-in-distributed-range mechanism with Shannon entropy providing the offsets. We will

[1]Here, we consider individual rationality in expectation. Achieving individual rationality in the *ex-post* sense is impossible for any non-trivial private mechanism since the probability of a non-zero price would have to jump by an infinitely large factor as an agent changes from zero valuation to non-zero valuation.

define affine maximum-in-distributed-range mechanisms in Section II and more details on this observation are deferred to Section III-A. Readers are referred to [8]–[11] for recent applications of maximum-in-distributed-range mechanisms in algorithmic mechanism design.

Our proof is by connecting the exponential mechanism to the Gibbs measure and free energy in statistical mechanics. We exploit this connection to provide a simple proof of the incentive compatibility of the mechanism. We believe this intriguing connection is of independent interest and may lead to new ways of understanding the exponential mechanism and differential privacy.

While we do not have an efficient way of computing the allocation and prices of the exponential mechanism in general (this is also not known for VCG), we do show that in special cases such as multi-item auctions and procurement auctions for spanning tree, we can efficiently implement the exponential mechanism either exactly or approximately. Further, we show that the trade-off between privacy and social welfare in the exponential mechanism is asymptotically optimal in these two cases, even if we compare to mechanisms that need not be truthful. We also include another application of the exponential mechanism for the combinatorial public project problem where the social welfare is close to optimal for an arbitrarily small constant ϵ.

Interestingly, our implementation of the exponential mechanism for multi-item auctions has further implications in the recent work on blackbox reductions in Bayesian mechanism design [3], [17]. Combining our exponential mechanism for the matching market with the blackbox reduction procedure in [3], [17], we can get a blackbox reduction that converts any algorithm into BIC, differentially private mechanisms. We will leave further discussions to the relevant section.

B. Related Work

McSherry and Talwar [22] first proposed using differentially private mechanisms to design auctions by pointing out that differential privacy implies approximate incentive compatibility as well as resilience to collusion. In particular, they study the problem of revenue maximization in digital auctions and attribute auctions. They propose the exponential mechanism as a solution for these problems. McSherry and Talwar also suggest using the exponential mechanism to solve mechanism design problems with different objectives, such as social welfare.[2] Their instantiation of the exponential mechanism is differentially private, but only approximately truthful. Nissim et al. [24] show how to convert differentially private mechanisms into exactly truthful mechanism in some settings. However, the mechanism loses its privacy property after such conversion. Xiao [28] seeks to design mechanisms that are both differentially private and perfectly truthful and proposes a method to convert any truthful mechanism into a differentially private and truthful one when the type space is small. Unfortunately, it does not seem possible to extend the results in [24], [28] to more general mechanism design problems, while our result applies to *any* mechanism design problem (with payments).

Xiao [28] also proposed to explicitly model the agents' concern for privacy in the utilities by assuming agent i has a disutility that depends on the amount of information ϵ_i leaked by the mechanism. Chen et al. [7] and Nissim et al. [25] explored this direction and introduced truthful mechanisms for some specific problems. Exact evaluation of an agent's dis-utility usually requires knowledge of the types of all agents and hence this kind of mechanism can only be private if agents do not need to exactly compute their own dis-utility. The above works circumvent this issue by designing strictly truthful and sufficiently private mechanisms such that any agent's gain in privacy by lying is outweighed by the loss in the usual notion of utility, regardless of the exact value of dis-utility for privacy.

Finally, Ghosh and Roth [16] study the problem of selling privacy in auctions, which can be viewed as an orthogonal approach to combining mechanism design and differential privacy.

II. Preliminaries

A mechanism design problem is defined by a set of n agents and a range R of feasible outcomes. Throughout this paper we will assume the range R to be discrete, but all our results can be easily extended to continuous ranges with appropriate integrability. Each agent i has a private valuation $v_i : R \mapsto [0, 1]$. A central entity chooses one of the outcomes based on the agents' (reported) valuations. We will let $\mathbf{0}$ denote the all-zero valuation and let v_{-i} denote the valuations of every agent except i.

For the sake of presentation, we will assume that the agents' valuations can be any functions mapping the range of feasible outcomes to the interval $[0, 1]$. It is worth noting that since our mechanisms are incentive compatible in this setting, they are also automatically incentive compatible for more restricted valuations (e.g., submodular valuations for a combinatorial public project problem).

A *mechanism* M consists of an allocation rule $x(\cdot)$ and a payment rule $p(\cdot)$. The mechanism first lets the agents submit their valuations. However, an agent may strategically submit a false valuation if that is beneficial to her. We will let $b_1, \ldots, b_n : R \mapsto [0, 1]$ denote the *reported valuations* (bids) from the agents and let $\boldsymbol{b}$ denote the vector of these valuations. After the agents submit their bids, the allocation rule $x(\cdot)$ chooses a feasible outcome $r = x(\boldsymbol{b}) \in R$ and the payment rule $p(\cdot)$ chooses a vector of payments $p(\boldsymbol{b}) \in \mathbb{R}^n$. We will let $p_i(\boldsymbol{b})$ denote the payment for agent i. Note that both $x(\cdot)$ and $p(\cdot)$ may be randomized. We will consider the

[2]The main difference between our instantiation of the exponential mechanism and that by McSherry and Talwar is that we use properly chosen payments to incentivize agents to report truthfully.

standard setting of quasi-linear utility: given the allocation rule, the payment rule, and the reported valuations $\boldsymbol{b}$, for each $i \in [n]$, the *utility* of agent i is

$$u_i(v_i, x(\boldsymbol{b}), p_i(\boldsymbol{b})) = v_i(x(\boldsymbol{b})) - p_i(\boldsymbol{b}) \ .$$

We will assume the agents are risk-neutral and aim to maximize their expected utilities.

The goal is to design polynomial time mechanisms M that satisfy various objectives. In this paper, we will focus on the problem of maximizing the expected *social welfare*, which is defined to be the sum of the agents' valuations: $\mathbf{E}[\sum_{i=1}^{n} v_i(x(\boldsymbol{b}))]$.

Besides the expected social welfare, we take into consideration the strategic play of utility-maximizing agents and their concern about the mechanism leaking non-trivial information about their private data. Thus, we will restrict our attention to mechanisms that satisfy several game-theoretic requirements and have a privacy guarantee that we will define in the rest of this section.

A. Game-Theoretical Solution Concepts

A mechanism is *incentive compatible* (IC) if truth-telling is a dominant strategy, i.e., by reporting the true values an agent always maximizes her expected utility regardless of what other agents do - $v_i \in \arg\max_{b_i} \mathbf{E}[v_i(x(b_i, b_{-i})) - p_i(b_i, b_{-i})]$. We will also consider an approximate notion of truthfulness. A mechanism is γ-incentive compatible (γ-IC) if no agent can get more than γ extra utility by lying. Further, a mechanism is *individually rational* (IR) if the *expected* utility of each agent is always non-negative, assuming this agent reports truthfully: $\mathbf{E}[v_i(x(v_i, b_{-i})) - p_i(v_i, b_{-i})] \geq 0$. We seek to design mechanisms that are incentive compatible and individually rational.

Affine Maximum-In-Distributed-Range: An allocation rule $x(\cdot)$ is an *affine maximum-in-distributed-range allocation* if there is a set S of distributions over feasible outcomes, parameters $a_1, \ldots, a_n \in \mathbb{R}^+$, and an offset function $c : S \mapsto \mathbb{R}$, such that the $x(v_1, \ldots, v_n)$ always chooses the distribution $\nu \in S$ that maximizes

$$\mathbf{E}_{r \sim \nu}\left[\sum_{i=1}^{n} a_i v_i(r)\right] + c(\nu) \ .$$

In this paper, we are particularly interested in the case when $a_i = 1$, $\forall i \in [n]$, and c is the Shannon entropy of the distribution scaled by an appropriate parameter.

The affine maximum-in-distributed-range mechanisms can be interpreted as slight generalizations of the well-studied maximum-in-distributed-range mechanisms. If $a_i = 1$ for every $i \in [n]$ and $c(\cdot) = 0$, then such allocation rules are referred to as maximum-in-distributed-range (MIDR) allocations. There are well-known techniques for charging proper prices to make MIDR allocations and their affine generalizations incentive compatible. The resulting mechanisms are called MIDR mechanisms. MIDR mechanisms are important tools for designing computationally efficient mechanisms that are incentive compatible and approximate social welfare well (e.g., see [8]–[11]).

B. Differential Privacy

Differential privacy is a notion of privacy that has been studied the most in the theoretical computer science community over the past decade. It requires the distribution of outcomes to be nearly identical when the agent profiles are nearly identical. Formally,

Definition 1. *A mechanism is* ϵ-differentially private *if for any two valuation profiles* $\boldsymbol{v} = (v_1, \ldots, v_n)$ *and* $\boldsymbol{v}' = (v'_1, \ldots, v'_n)$ *such that only one agent has different valuations in the two profiles, and for any set of outcomes* $S \subseteq R$*, we have*

$$\mathbf{Pr}[x(\boldsymbol{v}) \in S] \leq \exp(\epsilon) \cdot \mathbf{Pr}[x(\boldsymbol{v}') \in S] \ .$$

This definition of privacy has many appealing theoretical properties. Readers are referred to [13], [14] for excellent surveys on the subject.

We will also consider a standard variant that defines a more relaxed notion of privacy.

Definition 2. *A mechanism is* (ϵ, δ)-differentially private *if for any two valuation profile* $\boldsymbol{v} = (v_1, \ldots, v_n)$ *and* $\boldsymbol{v}' = (v'_1, \ldots, v'_n)$ *such that only one agent has different valuations in the two profiles, and for any set of outcomes* $S \subseteq R$*,*

$$\mathbf{Pr}[x(\boldsymbol{v}) \in S] \leq \exp(\epsilon) \cdot \mathbf{Pr}[x(\boldsymbol{v}') \in S] + \delta \ .$$

Typically, we will consider very small values of δ, say, $\delta = \exp(-n)$.

Differentially Private Payment: In the above definitions, we only consider the privacy of the allocation rule. We note that in practice, the payments need to be differentially private as well. We can handle privacy issues in the payments by the standard technique of adding Laplace noise. In particular, if the payments are implemented via secure channels (e.g. the same channels that the agents use to submit their bids) such that the each agent's payment is accessible only by the agent herself and the central entity, then adding independent Laplace noise with standard deviation $O(\epsilon^{-1})$ is sufficient to guarantee ϵ-differentially private payments. Since the techniques used to handle payments are quite standard, we will defer the extended discussion of this subject to the appendix.

C. The Exponential Mechanism

One powerful tool in the differential privacy literature is the exponential mechanism of McSherry and Talwar [22]. The exponential mechanism is a general technique for constructing differentially private algorithms over an arbitrary range R of outcomes and any objective function $Q(D, r)$ (often referred to as the quality function in the differential privacy literature) that maps a pair consisting

of a data set D and a feasible outcome $r \in R$ to a real-valued score. In our setting, D is a (reported) valuation profile and the quality function $Q(\boldsymbol{v}, r) = \sum_{i=1}^{n} v_i(r)$ is the social welfare.

Given a range R, a data set D, a quality function Q, and a privacy parameter ϵ, the *exponential mechanism* $\textsc{Exp}(R, D, Q, \epsilon)$ chooses an outcome r from the range R with probability

$$\mathbf{Pr}\left[\textsc{Exp}(R, D, Q, \epsilon) = r\right] \propto \exp\left(\frac{\epsilon}{2\Delta} Q(D, r)\right) \ ,$$

where Δ is the Lipschitz constant of the quality function Q, that is, for any two adjacent data set D_1 and D_2, and for any outcome r, the score $Q(D_1, r)$ and $Q(D_2, r)$ differs by at most Δ. In out setting, the Lipschitz constant of the social welfare function is 1. We sometimes use $\textsc{Exp}(D, \epsilon)$ for short when the range R and the quality function Q is clear from the context. We will use the following theorem about the exponential mechanism.

Theorem 1 (e.g., [22], [27])**.** *The exponential mechanism is ϵ-differentially private and ensures that*

$$\mathbf{Pr}\left[Q(D, \textsc{Exp}(D, \epsilon)) < \max_{r \in R} Q(D, r) - \frac{\ln |R|}{\epsilon} - \frac{t}{\epsilon}\right] \leq \exp(-t) \ .$$

III. The Exponential Mechanism is Incentive Compatible

In this section, we will show that if we choose the social welfare to be the quality function, then the exponential mechanism can be implemented in an incentive compatible and individually rational manner. Formally, for any range R and any privacy parameter $\epsilon > 0$, the exponential mechanism $\textsc{Exp}_{\epsilon}^{R}$ with its pricing scheme is presented in Figure 1. Our main theorem is the following:

Theorem 2. *The exponential mechanism with our pricing scheme is IC and IR.*

Our proof of Theorem 2 relies on the connection between the exponential mechanism and a well known probability measure in probability and statistical mechanics called the Gibbs measure. Once we have established this connection, the proof of Theorem 2 becomes very simple.

A. The Exponential Mechanism and the Gibbs Measure

The Gibbs measure, also known as the Boltzmann distribution in chemistry and physics, is formally defined as follows:

Definition 3 (Gibbs measure)**.** *Suppose we have a system consisting of particles of a gas. If the particles have k states $1, \ldots, k$, possessing energy $E_1, \ldots, E_k$ respectively, then the probability that a random particle in the system has state i follows the* Gibbs measure*:*

$$\mathbf{Pr}[\textit{state} = i] \propto \exp\left(-\tfrac{1}{k_B T} E_i\right) \ ,$$

where T is the temperature, and k_B is the Boltzmann constant.

Note that the Gibbs measure asserts that nature prefers states with lower energy level. Indeed, if $T \to 0$, then almost surely we will see a particle with lowest-energy state. On the other hand, if $T \to +\infty$, then all states are equally likely to appear. Thus the temperature T is a measure of uncertainty in the system: the lower the temperature, the less uncertainty in the system, and vice versa.

Gibbs Measure vs. Exponential Mechanism: It is not difficult to see the analogy between the Gibbs measure and the exponential mechanism. Firstly, the quality $Q(r)$ of an outcome $r \in R$ (in our instantiation, $Q(r)$ is the social welfare $\sum_i v_i(r)$) is an analog of the energy (more precisely, the negative of the energy) of a state i. In the exponential mechanism the goal is to maximize the expected quality of the outcome, while in physics nature tries to minimize the expected energy. Second, the privacy parameter ϵ is an analogue of the inverse temperature T^{-1}, both measuring the level of uncertainty in the system. The more privacy we want in the mechanism, the more uncertainty we need to impose in the distribution of outcomes[3]. Finally, the Lipschitz constant Δ and Boltzmann constant k_B are both scaling factors that come from the environment. Table I summarize this connection between the Gibbs measure and the exponential mechanism.

Gibbs Measure Minimizes Free Energy: It is well-known that the Gibbs measure maximizes entropy given the expected energy. In fact, a slightly stronger claim (e.g. see [21]) states that the Gibbs measure minimizes free energy. To be precise, suppose T is the temperature, ν is a distribution over the states, and $S(\nu)$ is the Shannon entropy of ν. Then the *free energy* of the system is

$$F(\nu, T) = \mathbf{E}_{i \sim \nu}[E_i] - k_B T \cdot S(\nu) \ .$$

The following result is well known in the statistical physics literature.

Theorem 3 (e.g. see [21])**.** *$F(\nu, T)$ is minimized when ν is the Gibbs measure.*

For self-containedness, we include the proof of Theorem 3 as follows.

Proof: Note that the free energy can be written as

$$\begin{aligned} & F(\nu, T) \\ = \ & \mathop{\mathbf{E}}_{i \sim \nu}[E_i] - k_B T \cdot S(\nu) \\ = \ & \sum_i \mathop{\mathbf{Pr}}_{\nu}[i] E_i + k_B T \sum_i \mathop{\mathbf{Pr}}_{\nu}[i] \ln \mathop{\mathbf{Pr}}_{\nu}[i] \ . \end{aligned} \tag{1}$$

[3]We note that the privacy guarantee ϵ is not necessarily a monotone function of the entropy of the outcome distribution. So the statement above is only for the purpose of establishing a high-level connection between the Gibbs measure and the exponential mechanism.

1) Choose outcome $r \in R$ with probability $\mathbf{Pr}[r] \propto \exp\left(\frac{\epsilon}{2}\sum_i v_i(r)\right)$.
2) For $1 \le i \le n$, charge agent i price

$$p_i = -\mathop{\mathbf{E}}_{r\sim \textsc{Exp}_\epsilon^R(b_i,b_{-i})}\left[\sum_{k\neq i} b_k(r)\right] - \frac{2}{\epsilon}\cdot S\left(\textsc{Exp}_\epsilon^R(b_i,b_{-i})\right) + \frac{2}{\epsilon}\ln\left(\sum_{r\in R}\exp\left(\frac{\epsilon}{2}\sum_{k\neq i} v_k(r)\right)\right) \ ,$$

where $S(\cdot)$ is the Shannon entropy.

Figure 1. $\textsc{Exp}_\epsilon^R$: the incentive-compatible exponential mechanism.

Table I
A HIGH-LEVEL COMPARISON BETWEEN THE GIBBS MEASURE AND THE EXPONENTIAL MECHANISM

	Gibbs measure	Exponential mechanism
Probability mass function	$\mathbf{Pr}[\text{state} = i] \propto \exp\left(-\frac{1}{k_B T}E_i\right)$	$\mathbf{Pr}[\text{outcome} = r] \propto \exp\left(\frac{\epsilon}{2\Delta}Q(r)\right)$
Objective function	$-E_i$	$Q(r)$
Measure of uncertainty	temperature T	privacy parameter ϵ
Environment parameter	Boltzmann constant k_B	Lipschitz constant Δ

Further, the first term of the right hand side can be rewritten as

$$\begin{aligned}
&\sum_i \mathop{\mathbf{Pr}}_\nu[i]E_i \\
=\ & k_B T\sum_i \mathop{\mathbf{Pr}}_\nu[i]\frac{1}{k_B T}E_i \\
=\ & -k_B T\sum_i \mathop{\mathbf{Pr}}_\nu[i]\ln\left(\exp\left(-\frac{1}{k_B T}E_i\right)\right) \\
=\ & -k_B T\sum_i \mathop{\mathbf{Pr}}_\nu[i]\ln\left(\frac{\exp\left(-\frac{1}{k_B T}E_i\right)}{\sum_j \exp\left(-\frac{1}{k_B T}E_j\right)}\right) \\
& -k_B T\ln\left(\sum_j \exp\left(-\frac{1}{k_B T}E_j\right)\right) \\
=\ & -k_B T\sum_i \mathop{\mathbf{Pr}}_\nu[i]\ln\left(\mathop{\mathbf{Pr}}_{\text{Gibbs}}[i]\right) \\
& -k_B T\ln\left(\sum_j \exp\left(-\frac{1}{k_B T}E_j\right)\right) \quad . \qquad (2)
\end{aligned}$$

By (1) and (2), the free energy equals

$$\begin{aligned}
F(\nu,T) = {} & k_B T\cdot D_{KL}(\nu\,||\,\text{Gibbs}) \\
& - k_B T\ln\left(\sum_j \exp\left(-\frac{1}{k_B T}E_j\right)\right) \quad .
\end{aligned}$$

Note that the second term is independent of ν. By basic properties of the KL-divergence, the above is minimized when ν is the Gibbs measure. ■

B. Proof of Theorem 2

By the connection between Gibbs measure and exponential mechanism and Theorem 3, we have the following analogous lemma for our instantiation of the exponential mechanism.

Lemma 1. *The* **free social welfare**,

$$\mathbf{E}_{r\sim\nu}\left[\textstyle\sum_i v_i(r)\right] + \frac{2}{\epsilon}\cdot S(\nu) \ ,$$

is maximized when $\nu = \textsc{Exp}_\epsilon^R(v_1,\ldots,v_n)$.

Incentive Compatibility: Let us consider a particular agent i, and fix the bids b_{-i} of the other agents. Suppose agent i has value v_i and bids b_i. For notational convenience, we let $b(r) = \sum_{k=1}^n b_k(r)$ and let

$$h_i(b_{-i}) = \frac{2}{\epsilon}\ln\left(\sum_{r\in R}\exp\left(\frac{\epsilon}{2}\sum_{k\neq i} v_k(r)\right)\right) \quad .$$

Using the price p_i charged to agent i as in Figure 1, her utility when she bids b_i is

$$\begin{aligned}
&\mathop{\mathbf{E}}_{r\sim\textsc{Exp}_\epsilon^R(b_i,b_{-i})}[v_i(r) + \sum_{k\neq i} b_k(r)] \\
&+ \frac{2}{\epsilon}\cdot S(\textsc{Exp}_\epsilon^R(b_i,b_{-i})) - h_i(b_{-i}) \ ,
\end{aligned}$$

which equals the free social welfare plus a term that does not depend on agent i's bid. By Lemma 1, the free social welfare is maximized when we use the outcome distribution by the exponential mechanism with respect to agent i's true value. Therefore, truthful bidding is a utility-maximizing strategy for agent i.

Individual Rationality: We first note that for any agent i, it is not difficult to verify that $p_i = 0$ when $v_i = \mathbf{0}$ regardless of bidding valuations of other agents. Therefore, by bidding $\mathbf{0}$ agent i could always guarantee non-negative expected utility. Since we have shown that the exponential mechanism is truthful-in-expectation, we get that the utility

of agent i when she truthfully reports her valuation is always non-negative.

Remark 1. We notice that Lemma 1 implies that the allocation rule of the exponential mechanism is affine maximum-in-distributed-range. As a result, there are standard techniques to charge prices so that the mechanisms is IC and IR as presented above.

Remark 2. Alternatively, one can prove Theorem 2 via the procedure developed by Rochet [26]: first prove the cyclic monotonicity of the exponential allocation rule, which is known to be the necessary and sufficient condition for being the allocation rule of a truthful mechanism; then derive the pricing scheme that rationalizes the exponential allocation rule via Rochet's characterization. We will omit further details of this proof in this extended abstract.

IV. Generalization

In the original definition by McSherry and Talwar [22], the exponential mechanism is defined with respect to a prior distribution $\mu(\cdot)$ over the feasible range R. More precisely, the exponential mechanism given μ, $\textsc{Exp}_\mu(R, D, Q, \epsilon)$, chooses an outcome r from the range R with probability

$$\mathbf{Pr}\left[\textsc{Exp}_\mu(R, D, Q, \epsilon) = r\right] \propto \mu(r) \exp\left(\frac{\epsilon}{2\Delta} Q(D, r)\right) \quad .$$

When μ is chosen to be the uniform distribution over the feasible range, we recover the definition in Section II. Using a different μ can improve computational efficiency as well as the trade-off between privacy and the objective for some problems (e.g., [5]). In every use of the (generalized) exponential mechanism, to our knowledge, μ is taken to be the uniform distribution over a sub-range that forms a geometric covering of the feasible range. But in general, this need not be the optimal choice.

We observe that our result can be extended to the above generalized exponential mechanism as well. More precisely, we can show that the generalized exponential mechanism is affine maximum-in-distributed-range as well.

Theorem 4. *For any range R, any quality function Q, any privacy parameter ϵ, any prior distribution μ, and any database D, the generalized exponential mechanism satisfies*

$$\textsc{Exp}_\mu(R, D, Q, \epsilon) = \arg\max_\nu \mathop{\mathbf{E}}_{r\sim\nu}[Q(D, r)] - \frac{2}{\epsilon} D_{KL}(\nu||\mu) \quad .$$

Corollary 5. *For any mechanism design problem for social welfare and any prior distribution μ over the feasible range, the generalized exponential mechanism (w.r.t. μ) is IC and IR with appropriate payment rule.*

The proof of Theorem 4 and deriving the pricing scheme in Corollary 5 is very similar to the corresponding parts in Section III and hence omitted.

V. Applications

Our result in Theorem 2 applies to a large family of problems. In fact, it can be used to derive truthful and differentially private mechanisms for any problem in mechanism design (with payments) that aims for social welfare maximization.

In this section, we will consider three examples – the combinatorial public project problem (CPPP), the multi-item auction, and the procurement auction for a spanning tree. The exponential mechanism for the combinatorial public project problem is incentive compatible, ϵ-differentially private, and achieves nearly optimal social welfare for any constant $\epsilon > 0$. However, we cannot implement the exponential mechanism in polynomial time for CPPP in general because implementing VCG for CPPP is known to be **NP**-hard and the exponential mechanism is a generalization of VCG. For the other two applications, we manage to implement the exponential mechanism in polynomial time, where the implementation for multi-item auction is only approximate so that it is only approximately truthful and approximately differentially private, and the implementation for procurement auction for spanning trees is exact. The social welfare for these two cases, however, is nearly optimal only when the privacy parameter ϵ is super-constantly large. Nonetheless, we show that the trade-offs between privacy and social welfare of the exponential mechanism in these two applications are asymptotically optimal.

A. Combinatorial Public Project Problem

The first interesting application of our result is a truthful and differentially private mechanism for the Combinatorial Public Project Problem (CPPP) originally proposed by Papadimitriou et al. [25]. In CPPP, there are n agents and m public projects. Each agent i has a private valuation function v_i that specifies agent i's value (between 0 and 1) for every subset of public projects. The objective is to find a subset S of public projects to build, of size at most k (a parameter), that maximizes the social welfare, namely, $\sum_i v_i(S)$.

This problem has received a lot of attention in the algorithmic game theory literature because strong lower bounds can be shown for the approximation ratio of this problem by any truthful mechanism when the valuations are submodular (e.g. see [12], [25]).

Further, the CPPP is of practical interest as well. The following is a typical CPPP scenario in the real world. Suppose some central entity (e.g. the government) wants to build several new hospitals where there are m potential locations to choose from. Due to resource constraints, the government can only build k hospitals. Each citizen has a private value for each subset of locations that may depend on the distance to the closest hospital and the citizen's health status.

Note that the agents may be concerned about their privacy if they choose to participate in the mechanism because

their valuations typically contain sensitive information. For example, the citizens who have high values for having a hospital close by in the above scenario are more likely to have health problems. Therefore, it would be interesting to design mechanisms for the CPPP that are not only truthful but also differentially private. The size of the range of outcomes is $\binom{m}{k} = O(m^k)$. So by Theorem 1 and Theorem 2, we have the following.

Theorem 6. *For any $\epsilon > 0$, the exponential mechanism $\text{EXP}_\epsilon^{\text{CPPP}}$ for CPPP is IC, ϵ-differentially private, and ensures*

$$\mathbf{Pr}\left[\sum_{i=1}^{n} v_i\left(\text{EXP}_\epsilon^{\text{CPPP}}\right) < \mathsf{opt} - \frac{k\ln m}{\epsilon} - \frac{t}{\epsilon}\right] \le \exp(-t) \ .$$

It is known that the exponential mechanism achieves the optimal trade-off between privacy and social welfare for CPPP (e.g., [27]).

Further, note that the optimal social welfare could be as large as n. Moreover, the number of projects $k \le m$ is typically much smaller than the number of agents n. Therefore, the exponential mechanism achieves social welfare that is close to optimal. However, it is worth noting that we only requires k and m to be mildly smaller than n (e.g. $O(n^{1-c})$ for any small constant $c > 0$), in which cases the size of the type space, which is exponential in k and m, is still quite large so that the approach in [28] does not apply.

In some scenarios such as the one above where the government wants to build a few new hospitals, k is sufficiently small so that it is acceptable to have running time polynomial in the size of the range of outcomes. In such cases, it is easy to see that the exponential mechanism for CPPP can be implemented in time polynomial in n and $\binom{m}{k}$.

B. Multi-Item Auction

Next we consider a multi-item auction. Here, the auctioneer has n heterogeneous items (one copy of each item) that she wishes to allocate to n different agents[4]. Agent i has a private valuation $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{ik})$, where v_{ij} is her value for item j. We will assume the agents are unit-demand, that is, each agent wants at most one item. It is easy to see that each feasible allocation of the multi-item auction is a matching between agents and items. We will let the R_M denote the range of multi-item auction, that is, the set Π_n of all permutations on $[n]$.

The multi-item auction and related problems are very well-studied in the algorithmic game theory literature (e.g. [4], [7]). They capture the motivating scenario of allocating oil fields and many other problems that arise from allocating public resources. The VCG mechanism can be implemented in polynomial time to maximize social welfare in this problem since max-matching can be solved in polynomial time. The new twist in our setting is to design mechanisms that are *both truthful and differentially private* and have good social welfare guarantee.

Approximate Implementation of the Exponential Mechanism: Unfortunately, exactly sampling matchings according to the distribution specified in the exponential mechanism seems hard due to its connection to the problem of computing the permanent of non-negative matrices (e.g. see [18]), which is $\#P$-complete. Instead, we will sample from the desired distribution approximately. Moreover, we show that there is an efficient approximate implementation of the payment scheme. As a result of the non-exact implementation, we only get γ-IC instead of perfect IC, (ϵ, δ)-differential privacy instead of ϵ-differential privacy, and lose an additional $n\gamma$ additive factor in social welfare. Here, γ will be inverse polynomially small. The discussion of this approximate implementation of the exponential mechanism is deferred to the full version.

Note that the size of the range of feasible outcomes of multi-item auction is $n!$. By Theorem 1, we have the following:

Theorem 7. *For any $\delta \in (0,1)$, $\epsilon > 0$, $\gamma > 0$, there is a polynomial time (in n, ϵ^{-1}, γ^{-1}, and $\log(\delta^{-1})$) approximate implementation of the exponential mechanism, $\widehat{\text{EXP}}_\epsilon^{R_M}$ that is γ-IC, (ϵ, δ)-differentially private, and ensures that*

$$\mathbf{Pr}\left[\sum_{i=1}^{n} v_i\left(\widehat{\text{EXP}}_\epsilon^{R_M}\right) < \mathsf{opt} - \gamma n - \frac{\ln(n!)}{\epsilon} - \frac{t}{\epsilon}\right] \le \exp(-t) \ .$$

Note that here we are achieving γ-IC and (ϵ, δ)-differentially privacy while in the instantiation of the exponential mechanism by McSherry and Talwar [22] is ϵ-IC and ϵ-differentially private. Our result in Theorem 7 is better in most applications since typically ϵ is large, usually a constant or occasionally a super-constant, while γ is small, usually requires to be $1/\text{poly}$ for γ-IC to be an appealing solution concept.

The trade-off between privacy and social welfare in Theorem 7 can be interpreted as the follows: if we want to achieve social welfare that is worse than optimal by at most an $O(n)$ additive term, then we need to choose $\epsilon = \Omega(\log n)$. The next theorem shows that this is tight. The proof is deferred to the full version.

Theorem 8. *Suppose M is an ϵ-differentially private mechanism for the multi-item auction problem and the expected welfare achieve by M is at least $\mathsf{opt} - \frac{n}{10}$. Then $\epsilon = \Omega(\log n)$.*

Note that in this theorem, we do not restrict M to be incentive compatible. In other word, this lower bound holds for arbitrary differentially private mechanisms. So there is no extra cost for imposing the truthfulness constraint.

[4]The case when the number of items is not the same as the number of agents can be reduced to this case by adding dummy items or dummy agents. So our setting is w.l.o.g.

Implication in BIC Blackbox Reduction: Recently, Hartline et al. [17] and Bei and Huang [3] introduce blackbox reductions that convert any algorithm into nearly Bayesian incentive-compatible mechanisms with only a marginal loss in the social welfare. Both approach essentially create a virtual interface for each agent which has the structure of a matching market and then run VCG in the virtual matching markets. By running the exponential mechanism instead of the VCG mechanism, we can obtain a blackbox reduction that converts any algorithm into a nearly Bayesian incentive-compatible and differentially private mechanism. We will defer more details to the full version of this paper.

C. Procurement Auction for Spanning Trees

Another interesting application is the procurement auction for a spanning tree (e.g. see [6]). Procurement auctions (also known as reverse auctions) are a type of auction where the roles of buyers and sellers are reversed. In other word, the central entity seeks to buy, instead of sell, items or services from the agents. In particular in the procurement auction for spanning trees, consider $n = \binom{k}{2}$ selfish agents own edges in a publicly known network of k nodes. We shall imagine the nodes to be cities and the edges as potential highways connecting cities. Each agent i has a non-negative cost c_i for building a highway along the corresponding edge. The central entity (e.g. the government) wants to purchase a spanning tree from the network so that she can build highways to connect the cities. The goal is to design incentive compatible and differentially private mechanisms that provide good social welfare (minimizing total cost).

Although this is a reverse auction in which agents have costs instead of values and the payments are from the central entity to the agents, by interpreting the costs as the negative of the valuations (i.e. $v_i = -c_i$ if the edge is purchased and $v_i = 0$ otherwise), we can show that the exponential mechanism with the same payment scheme is incentive compatible for procurement auctions via almost identical proofs. We will omit the details in this extended abstract.

Next, we will discuss how to efficiently implement the exponential mechanism.

Sampling Spanning Trees: There has been a large body of literature on sampling spanning tree (e.g. see [20] and the reference therein). Recently, Asadpour et al. [1] have developed a polynomial time algorithm for sampling *entropy-maximizing* distributions, which is exactly the kind of distribution used by the exponential mechanism. Therefore, the allocation rule of the exponential mechanism can be implemented in polynomial time for the spanning tree auction.

Implicit Payment Scheme by Babaioff, Kleinberg, and Slivkins [2]: Although we can efficiently generate samples from the desired distribution, it is not clear how to compute the exact payment explicitly. Fortunately, Babaioff et al. [2], [19] provide a general method of computing an unbiased estimator for the payment given any rationalizable allocation rule[5]. Hence, we can use the implicit payment method in [2], [19] to generate the payments in polynomial time.

Note that the size of the range of feasible outcomes of spanning tree auction is the number of different spanning tree in a complete graph with k vertices, which equals k^{k-2}. By Theorem 1 we have the following:

Theorem 9. *For any $\epsilon > 0$, the exponential mechanism $\text{EXP}_\epsilon^{tree}$ runs in polynomial time (in k and ϵ^{-1}), is IC, ϵ-differentially private, and ensures that*

$$\mathbf{Pr}\left[\sum_{i=1}^{n} c_i\left(\widehat{\text{EXP}}_\epsilon^{tree}\right) > \mathsf{opt} + \frac{(k-2)\log k}{\epsilon} + \frac{t}{\epsilon}\right] \leq \exp(-t) \ .$$

This trade-off between privacy and social welfare in Theorem 9 essentially means that we need $\epsilon = \Omega(\log k)$ in order to get $\mathsf{opt} + O(k)$ guarantee on expected total cost. The next theorem shows that this tradeoff is also tight. The proof is deferred to the full version due to space constraint.

Theorem 10. *Suppose M is an ϵ-differentially private mechanism for the procurement auction for spanning tree and the expected total cost by M is at most $\mathsf{opt} + \frac{k}{24}$. Then $\epsilon = \Omega(\log k)$.*

Similar to the case in the multi-item auction, the above lower bound does not restrict M to be incentive compatible. So the exponential mechanism is optimal even if we compare it to non-truthful ones.

ACKNOWLEDGEMENT

The authors would like to thank Aaron Roth for many useful comments and helpful discussions.

REFERENCES

[1] A. Asadpour, M.X. Goemans, A. Madry, S.O. Gharan, and A. Saberi. An $O(\log n/\log\log n)$-approximation algorithm for the asymmetric traveling salesman problem. In *SODA*, pages 379–389. ACM-SIAM, 2010.

[2] M. Babaioff, R.D. Kleinberg, and A. Slivkins. Truthful mechanisms with implicit payment computation. In *EC*, pages 43–52. ACM, 2010.

[3] X. Bei and Z. Huang. Bayesian incentive compatibility via fractional assignments. In *SODA*. ACM-SIAM, 2011.

[4] S. Bhattacharya, G. Goel, S. Gollapudi, and K. Munagala. Budget constrained auctions with heterogeneous items. In *STOC*, pages 379–388. ACM, 2010.

[5] A. Blum, K. Ligett, and A. Roth. A learning theory approach to non-interactive database privacy. In *STOC*, pages 609–618. ACM, 2008.

[5] Although the result in [2] only applies to single-parameter problems, Kleinberg [19] pointed out the same approach can be extended to multi-parameter problems if the type space is convex.

[6] M.C. Cary, A.D. Flaxman, J.D. Hartline, and A.R. Karlin. Auctions for structured procurement. In *SODA*, pages 304–313. ACM-SIAM, 2008.

[7] S. Chawla, J. Hartline, D. Malec, and B. Sivan. Sequential posted pricing and multi-parameter mechanism design. In *STOC*, pages 311–320. ACM, 2010.

[8] S. Dobzinski and S. Dughmi. On the power of randomization in algorithmic mechanism design. In *FOCS*, pages 505–514. IEEE, 2009.

[9] S. Dughmi. A truthful randomized mechanism for combinatorial public projects via convex optimization. In *EC*, pages 263–272. ACM, 2011.

[10] S. Dughmi and T. Roughgarden. Black-box randomized reductions in algorithmic mechanism design. In *FOCS*, pages 775–784. IEEE, 2010.

[11] S. Dughmi, T. Roughgarden, and Q. Yan. From convex optimization to randomized mechanisms: toward optimal combinatorial auctions. In *STOC*, pages 149–158. ACM, 2011.

[12] S. Dughmi and J. Vondrak. Limitations of randomized mechanisms for combinatorial auctions. In *FOCS*, pages 502–511. IEEE, 2011.

[13] C. Dwork. A firm foundation for private data analysis. *Communications of the ACM, to appear*.

[14] C. Dwork. Differential privacy: A survey of results. In *TAMC*, pages 1–19, 2008.

[15] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. *Theory of Cryptography*, pages 265–284, 2006.

[16] A. Ghosh and A. Roth. Selling privacy at auction. In *EC*, pages 199–208. ACM, 2011.

[17] J. Hartline, R. Kleinberg, and A. Malekian. Bayesian incentive compatibility via matchings. In *SODA*. ACM-SIAM, 2011.

[18] M. Jerrum and A. Sinclair. Approximating the permanent. *SIAM Journal on Computing*, 18:1149, 1989.

[19] R.D. Kleinberg. Personal communication.

[20] V.G. Kulkarni. Generating random combinatorial objects. *Journal of Algorithms*, 11(2):185–207, 1990.

[21] A. Le Ny. Introduction to (generalized) Gibbs measures. *Ensaios Matemáticos*, 15:1–126, 2008.

[22] F. McSherry and K. Talwar. Mechanism design via differential privacy. In *FOCS*, 2007.

[23] N. Nisan, T. Roughgarden, E. Tardos, and V.V. Vazirani. *Algorithmic game theory*. Cambridge University Press, 2007.

[24] K. Nissim, R. Smorodinsky, and M. Tennenholtz. Approximately optimal mechanism design via differential privacy. *ITCS*, 2012.

[25] C. Papadimitriou, M. Schapira, and Y. Singer. On the hardness of being truthful. In *FOCS*, pages 250–259. IEEE, 2008.

[26] J.C. Rochet. A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of Mathematical Economics*, 16(2):191–200, 1987.

[27] K. Talwar, A. Gupta, K. Ligett, F. McSherry, and A. Roth. Differentially private combinatorial optimization. In *SODA*. ACM-SIAM, 2010.

[28] D. Xiao. Is privacy compatible with truthfulness? In *Cryptology ePrint Technical Report, 2011/005*, 2011.

APPENDIX

In this section, we will discuss what is the amount of noise one needs to add to the payments in order to achieve ϵ-differential privacy. We will consider two different models depending on how the payments are implemented: the *public payment* model and the *private payment* model.

In the public payment model, the payments of the agents will become public information at the end of the auction, that is, the adversary who tries to learn the private valuations of the agents can see all the payments. Therefore, a payment scheme is ϵ-differentially private in the public payment model if and only if for any $i \in [n]$, any value profiles $\boldsymbol{v} = (v_1, \ldots, v_n)$ and $\boldsymbol{v}' = (v_1, \ldots, v_i', \ldots, v_n)$ that differ only in the valuation of agent i, and any possible payment profile $\boldsymbol{p}$, the probability

$$\begin{aligned} & \mathbf{Pr}[p_1(\boldsymbol{v}), \ldots, p_n(\boldsymbol{v}) = \boldsymbol{p}] \\ \leq \ & \exp(\epsilon)\, \mathbf{Pr}[p_1(\boldsymbol{v}'), \ldots, p_n(\boldsymbol{v}') = \boldsymbol{p}] \ . \end{aligned}$$

In the private payment model, we will assume the payments are implemented via secure channels such that the payment of each agent is only known to the corresponding agent and a few trusted parties, e.g. the central entity who runs the mechanism and/or the bank. Here, there are two cases based on what information the adversary can learn from the payments. If the adversary is not one of the agents, then by our assumption, he cannot see any of the payments and therefore cannot learn any information from the payments. If the adversary is one of the agents, then the only information of the payments that he will have access to is his own payment. Therefore, a payment scheme is ϵ-differentially private in the public payment model if and only if for any $i \neq j \in [n]$, any value profiles $\boldsymbol{v} = (v_1, \ldots, v_n)$ and $\boldsymbol{v}' = (v_1, \ldots, v_i', \ldots, v_n)$ that differ only in the valuation of agent i, and any possible payment p of agent j, the probability

$$\mathbf{Pr}[p_j(\boldsymbol{v}) = p] \leq \exp(\epsilon)\, \mathbf{Pr}[p_j(\boldsymbol{v}') = p] \ .$$

We will measure the amount of noise in the payments using L_2 norm, that is, we aim to minimize the total variance of the agents' payments in the worst-case: $\max_{\boldsymbol{v}} \sum_{i=1}^{n} \mathbf{Var}[p_i(\boldsymbol{v})]$.

Next, we will proceed to analyze the amount of noise needed in each of the two models. We will start with an upper bound on the sensitivity of each agent's payment as a function of the bids.

Lemma 2. *For any $i, j \in [n]$, and any value profiles $\boldsymbol{v} = (v_1, \ldots, v_n)$ and $\boldsymbol{v}' = (v_1, \ldots, v'_i, \ldots, v_n)$ that only differ in the valuation of agent i, we have $|p_j(\boldsymbol{v}) - p_j(\boldsymbol{v}')| \leq 1$.*

Proof: Note that by Theorem 2, the exponential mechanism is individual rational. It is also easy to see that it has no positive transfer for that otherwise the zero-value agent could gain by lying. So by our assumption that the agents' valuations are always between 0 and 1, we have $0 \leq p_j(\boldsymbol{v}), p_j(\boldsymbol{v}') \leq 1$. So Lemma 2 follows trivially. ■

In the public payment model, the mechanism has to reveal a vector of n real numbers (the payments) at the end of the auction, where each entry has sensitivity 1 by Lemma 2. Therefore, we can use the standard treatment for answering numerical queries, namely, adding independent Laplace noise $\textsc{Lap}(\frac{n}{\epsilon})$ to each entry, where $\textsc{Lap}(b)$ is the Laplace distribution with p.d.f. $f_{\textsc{Lap}(b)}(x) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right)$. More precisely, we can show the following theorem.

Theorem 11. *In the public payment model, the following payment scheme is ϵ-differentially private and has total variance $O(n^{3/2}\epsilon^{-1})$, while maintaining the IC and IR in expectation: let $p_1, \ldots, p_n$ be the payments specified in the exponential mechanism (Figure 1); let $x_1, \ldots, x_n$ be i.i.d. variables following the Laplace distribution $\textsc{Lap}(\frac{n}{\epsilon})$; use payment scheme $(p_1 + x_1, \ldots, p_n + x_n)$.*

The proof follows by standard analysis of the Laplace mechanism (e.g. see [15]). So we will omit the details in this extended abstract. It is worth mentioning that since the problem of designing payment scheme in the public payment model is a special case of answering n non-linear numerical queries, it may be possible to reduce the amount of noise by using more specialized scheme on a problem-by-problem basis. However, we feel this is less insightful than the other results we have in this paper, so we will focus on general mechanisms and payment schemes that work for all mechanism design problems.

Now let us turn to the private payment model. By our previous discussion, the mechanism only need to release at most one real number to each potential adversary in this model. So one may expect much less noise is needed in this model. Indeed, we could again use the standard treatment of adding Laplace noise, but this time it suffices to add independent Laplace noise $\textsc{Lap}(\frac{1}{\epsilon})$ to each entry.

Theorem 12. *In the private payment model, the following payment scheme is ϵ-differentially private and has total variance $O(\sqrt{n}\epsilon^{-1})$, while maintaining the IC and IR: in expectation: let $p_1, \ldots, p_n$ be the payments specified in the exponential mechanism (Figure 1); let $x_1, \ldots, x_n$ be i.i.d. variables following the Laplace distribution $\textsc{Lap}(\frac{1}{\epsilon})$; use payment scheme $(p_1 + x_1, \ldots, p_n + x_n)$.*