Applications of Secure Multiparty Computation
P. Laud and L. Kamm (Eds.)

doi:10.3233/978-1-61499-532-6-150

Chapter 8 Mechanism Design and Strong Truthfulness

Yiannis GIANNAKOPOULOS [a]
[a] *University of Oxford, UK*

Abstract. In this chapter we give a very brief overview of some fundamentals from mechanism design, the branch of game theory dealing with designing protocols to cope with agents' private incentives and selfish behavior. We also present recent results involving a new, extended utilities model that can incorporate *externalities*, such as malicious and spiteful behavior of the participating players. A new notion of *strong truthfulness* is proposed and analyzed. It is based on the principle of punishing players that lie. Due to this, strongly truthful mechanisms can serve as subcomponents in bigger mechanism protocols in order to boost truthfulness. The related solution concept equilibria are discussed and the power of the decomposability scheme is demonstrated by an application in the case of the well-known mechanism design problem of scheduling tasks to machines for minimizing the makespan.

Keywords. Mechanism design, incentive compatibility, strong truthfulness, VCG mechanisms

Introduction

Game theory [1] is a discipline that studies the strategic behavior of players that interact with each other and are completely rational and selfish: they only care about maximizing their own personal utilities. Extensive work has been done in this field over the last 60 years and various solution concepts and models have been proposed that try to explain and predict the outcome of such games. The branch of game theory called *mechanism design* (MD) was developed to have, in a way, a completely inverse goal: if we as algorithm designers want to enforce a certain outcome to such a system of interacting selfish entities, what are the game-playing rules we should impose? Therefore, it seems that MD and SMC share some very important common fundamental priors, that is, trying to optimize a joint objective while dealing with players that have incentives to manipulate our protocols. In fact, there has been a long and fruitful interaction between ideas from game theory and cryptography, see e.g. [2,3,4,5,6].

First, we give a short introduction to the basic ideas of MD, mostly using auctions as a working example, and discuss important MD paradigms, like the VCG mechanism and various optimization objectives, like maximizing the social welfare or the seller's revenue. Next, we devote the remaining chapter to discussing some exciting recent results from [7]. We believe that these results are particularly relevant to cryptographic considerations, as they involve a new, extended utilities model for MD settings which can incorporate *externalities*, such as malicious and spiteful behavior of the participating agents. A new notion of *strong* truthfulness is presented and analyzed. It is based on the principle of punishing players that lie. Due to this, strongly truthful mechanisms can serve as subcomponents in bigger mechanism protocols in order to boost truthfulness in settings with externalities and achieve a kind of externalities-resistant performance. This decomposability scheme is rather general and powerful, and we show how it can be also adapted to the case of the well-known MD problem of scheduling tasks to unrelated parallel machines in order to minimize the required execution time (makespan).

1. Foundations

In this section we are going to present some fundamental notions from MD, upon which the exposition and results of subsequent sections are going to be built. We deliberately choose to make this presentation based on a simple, single-item auction paradigm, in order to demonstrate more clearly both the intuition and the essence of these notions and not get lost in the technicalities of the general models which, after all, are not essential for the needs of this book. The reader, of course, can find many good references including more formal and general introductions to the fascinating area of algorithmic mechanism design (see e.g. [8,9]).

1.1. Basic Notions from Game Theory and Mechanism Design

Assume the following traditional MD setting, in particular a single-item auction scenario. We have n players (also called agents), each of whom is willing to pay t_i, $i = 1,2,\dots,m$, in order to get the item. These are called the agents' *types*, and $\mathbf{t} = (t_1,t_2,\dots,t_n)$ is the *type profile*. Let us assume that all these belong to some domain T, where in our particular auction setting it is a natural assumption to consider $T = \mathbb{R}_+$. This is *private* information of the players, who report it to the auctioneer in the form of *bids* b_i, $i = 1,2,\dots,m$. The reason we discriminate between types and bids is that the players, as we will see shortly, may have reason to lie about their true types and misreport some $b_i \neq t_i$. Given the input by the players, i.e. the bid profile $\mathbf{b} = (b_1,b_2,\dots,b_n)$, the auctioneer needs to decide who gets the item and how much she is going to pay for it.

More formally, a *mechanism* $\mathcal{M} = (\mathbf{a},\mathbf{p})$ consists of two vectors: an *allocation* vector $\mathbf{a} = \mathbf{a}(\mathbf{b}) = (a_1(\mathbf{b}),a_2(\mathbf{b}),\dots,a_n(\mathbf{b})) \in [0,1]^n$ and a *payment* vector $\mathbf{p} = \mathbf{p}(\mathbf{b}) = (p_1(\mathbf{b}),p_2(\mathbf{b}),\dots,p_n(\mathbf{b})) \in \mathbb{R}_+^n$. If our mechanism is deterministic, $a_i(\mathbf{b})$ is simply an indicator variable of the value 0 or 1, denoting whether player i wins the item or not. If we allow for randomized mechanisms, $a_i(\mathbf{b})$ denotes the probability of player i winning the item. In the latter case, we must make sure that $\sum_{i=1}^n a_i(\mathbf{b}) \leq 1$ for all $\mathbf{b} \in T^n$. Also, agent i will have to submit a payment of $p_i(\mathbf{b})$ to the mechanism. We define the *utility* of player i to be his total happiness after taking part in the auction, known as his *valuation* $v_i(\mathbf{a},t_i) = a_i \cdot t_i$, minus the payment p_i he has submitted. Formally, utility is defined as:

$$u_i(\mathbf{b}|t_i) = v_i(\mathbf{a}(\mathbf{b}),t_i) - p_i(\mathbf{b}) = a_i(\mathbf{b}) \cdot t_i - p_i(\mathbf{b}). \tag{1}$$

Notice the notation $u_i(\mathbf{b}|t_i)$ and the different usage of bids and types in expression (1). In case of *truth-telling*, i.e. honest reporting of $b_i = t_i$, we simplify the notation to

$$u_i(\mathbf{b}) = a_i(\mathbf{b}) \cdot b_i - p_i(\mathbf{b}). \tag{2}$$

We call these utilities *quasilinear*, due to the special form of these expressions, and in particular the linear-form connection between a player's utility and the player's own type t_i. The MD model can be defined in a more general way. The allocation function can belong to an arbitrary set of *outcomes* A. For our special case of single-item bidders, the outcomes are just the allocation vectors belonging to $A = [0,1]^n$. A player's valuation v_i is then defined over the set of possible outcomes and her true type t_i, $v_i(\mathbf{a}, t_i)$, and does not need to have the special linear form for the case of a single-item auction described previously. See Sec. 1.2.2 for a more general example of a MD setting.

If we look a little closer, we can see that we have already formed a *game* (see e.g. [10]): the players' strategies are exactly their valuations and each one of them is completely rational, and their goal is to *selfishly* maximize her own utility. So, we can use standard solution concepts and in particular *equilibria* in order to talk about possible stable states of our auctions. For example, the most fundamental notion that underlies the entire area of MD is that of *truthfulness* (also called incentive compatibility or strategy-proofness). Intuitively, we will say that a mechanism is truthful if it makes sure that no agent has an incentive to lie about her true type.

Definition 1 (Truthfulness)**.** *A mechanism $\mathcal{M} = (\mathbf{a}, \mathbf{p})$ is called truthful if truth-telling is a* dominant-strategy equilibrium *of the underlying game, i.e.* $u_i(b_i, \mathbf{b}_{-i}|b_i) \geq u_i(\tilde{b}_i, \mathbf{b}_{-i}|b_i)$ *for every player i, all possible bid profiles* $\mathbf{b} \in T^n$ *and all possible misreports* $\tilde{b}_i \in T$.

In case of randomized mechanisms, the above definitions are naturally extended by taking expectations of the utilities (*truthful in expectation* mechanisms). Here we have used standard game-theoretic notation, $\mathbf{b}_{-i}$ denoting the result of removing the i-th coordinate of $\mathbf{b}$. This vector is of one dimension lower than $\mathbf{b}$. This notation is very useful for modifying vectors at certain coordinates. For example $(x, \mathbf{b}_{-i})$ is the vector we get if we replace the i-th coordinate b_i of $\mathbf{b}$ with a new value of x. In particular, notice that this means that $(b_i, \mathbf{b}_{-i}) = \mathbf{b}$.

Being implemented in dominant strategies, truthfulness is a very stable and desirable property, which we want all our auction mechanisms to satisfy. It allows us to extract the truth from the participating parties and, thus, be able to accurately design the protocols for the goals we want to achieve. A celebrated result by Myerson gives us a powerful and simple characterization of truthful mechanisms and also helps us completely determine a mechanism simply by giving its allocation function $\mathbf{a}$.

Theorem 1 (Myerson [11])**.** *A mechanism $\mathcal{M} = (\mathbf{a}, p)$ is truthful if and only if:*

1. *Its allocation functions are monotone nondecreasing, in the sense that*

$$b_i \leq b_i' \quad \Longrightarrow \quad a_i(b_i, \mathbf{b}_{-i}) \leq a_i(b_i', \mathbf{b}_{-i})$$

for every player i, all valuation profiles $\mathbf{b}_{-i}$ *and all valuations* b_i, b_i'.

2. *The payment functions are given by*

$$p_i(\mathbf{b}) = a_i(\mathbf{b}) b_i - \int_0^{b_i} a_i(x, \mathbf{b}_{-i})\, dx. \tag{3}$$

Based on this, one can show another elegant and concise analytic characterization of truthful mechanisms.

Theorem 2. *[Rochet [12]] A mechanism is truthful if and only if for every player i, it induces a utility function u_i (over T^n) which is* convex *with respect to its i-th component.*

1.2. Fundamental MD Problems

In this section we briefly present two of the most fundamental MD domains, namely *additive auctions* (Sec. 1.2.1) and *scheduling unrelated machines* (Sec. 1.2.3), as well as the associated optimization objectives. These are the predominant motivating examples that move the entire field forward and also serve as the basis for our exposition in this chapter.

1.2.1. Welfare and Revenue in Auctions

The fundamental single-item auction introduced in Sec. 1.1 can be generalized to the following m-items *additive* valuations auction setting, where now the allocation $\mathbf{a}$ of a mechanism $\mathcal{M} = (\mathbf{a}, \mathbf{p})$ is an $n \times m$ matrix $\mathbf{a} = \{a_{ij}\} \subseteq [0,1]^{n \times m}$, where a_{ij} represents the probability of agent i getting item j, $i = 1,2,\dots,n$, $j = 1,2,\dots,m$. Inputs to the mechanism are bid profiles $\mathbf{b} \in T^{n \times m}$, $T = \mathbb{R}_+$, where b_{ij} is the bid of player i for item j. Of course, we must make sure that for every possible input $\mathbf{b}$ the selected outcome $\mathbf{a}(\mathbf{b})$ must not assign any item to more than one agent, i.e.

$$\sum_{i=1}^{n} a_{ij}(\mathbf{b}) \leq 1 \quad \text{for all items } j = 1,2,\dots,m. \tag{4}$$

When we design auction mechanisms we are usually interested in maximizing either the combined happiness of our society, meaning the sum of the valuations of the players that receive items, or the auctioneer's profit, i.e. the sum of the payments he collects from the participating agents. So, we define the social *welfare* of mechanism $\mathcal{M}$ on input $\mathbf{b}$ to be

$$W(\mathbf{b}) \equiv \sum_{i=1}^{n} v_i(\mathbf{a}(\mathbf{b}), \mathbf{b}_i) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}(\mathbf{b}) b_{ij} \ .$$

Here we assume that the players have *additive* valuations, that is, the valuation for receiving a subset of items is just the sum of the valuations of the items in the bundle, and its *revenue*

$$R(\mathbf{b}) \equiv \sum_{i=1}^{n} p_i(\mathbf{b}).$$

The most well-known auction is without doubt the VCG auction, named after the work of Vickrey [13], Clarke [14] and Groves [15]. In the simple single-item setting, this mechanism reduces to the Vickrey second-price auction that gives the item to the highest bidding agent but collects as payment the second highest bid. In that way, it ensures truthfulness by not giving an incentive to the wining agent to misreport a lower bid, as her payment is independent of her own bid and depends only on the bids of the other players. And, above all, this mechanism maximizes social welfare by definition.

Formally, by generalizing these ideas to the setting of m items, the VCG auction is a mechanism with the allocation rule

$$\mathbf{a}(\mathbf{b}) = \operatorname*{argmax}_{\alpha \in A} W(\alpha(\mathbf{b})) = \operatorname*{argmax}_{\alpha \in A} \sum_{i=1}^{n} v_i(\alpha, \mathbf{b}_i) \quad \text{for all } \mathbf{b} \in T^{n \times m}, \tag{5}$$

and payments

$$p_i(\mathbf{b}) = \max_{\alpha \in A} \sum_{j \neq i} v_j(\alpha(\mathbf{b}), \mathbf{b}_j) - \sum_{j \neq i} v_j(a(\mathbf{b}), \mathbf{b}_j) \quad \text{for all } \mathbf{b} \in T^{n \times m}. \tag{6}$$

The underlying idea is that first of all, the allocation rule (5) ensures that social welfare is the optimal one, and the payments (6) are such that they internalize the externalities of every player i. That is, intuitively, we charge player i for the harm that her participation in the auction causes to the rest of the society.

1.2.2. VCG Example: Buying Paths of a Network

We demonstrate the allocation and payments of the VCG mechanism through the following example, which is slightly more complex than the single-item auction we have been focusing on thus far. Let $G = (V, E)$ be a 2-edge connected directed graph. Let $s, t \in V$ be two nodes of the graph. Each edge $e \in E$ is a player that has as type an edge-cost (latency) c_e. We want to send a message from s to t, so the set of outcomes A in our setting are all possible paths π from s to t in G. We will denote such a path compactly as $\pi : s \to t$. If a player e is selected in an outcome-path, she incurs a damage of c_e, so the valuations are modeled by

$$v_e = v_e(\pi) = \begin{cases} -c_e, & \text{if } e \in \pi, \\ 0, & \text{otherwise.} \end{cases}$$

for all $\pi : s \to t$. Hence, maximizing social welfare $\sum_e v_e(\pi)$ in our problem is equivalent to selecting the shortest path from s to t in the weighted graph G with the edge weights c_e.

Consider the graph on Fig. 1. The outcome of our mechanism would be the shortest path from s to t, here $s \to a \to t$ that has the length of $5 + 1 = 6$. Let us look at what players (s, a) and (a, t) have to pay to get allocated. If player (s, a) was not present, the shortest path would be $s \to b \to t$ for social welfare of $(-4) + (-3) = -7$. As she is present, the welfare of the other players is $v_{(a,t)} + v_{(s,b)} + v_{(b,t)} = -1 + 0 + 0 = -1$. So, she should pay $p_{(s,a)} = -7 - (-1) = -6$. The negative sign means that we should give money to this edge/player for using it instead of asking money from it. Similarly, the monetary compensation to the other participating edge should be $p_{(a,t)} = -2$.

1.2.3. Scheduling and Minimizing Makespan

The scheduling domain is essentially an additive multi item auction setting (see Sec. 1.2.1) with the modification that players are not trying to maximize their utilities,

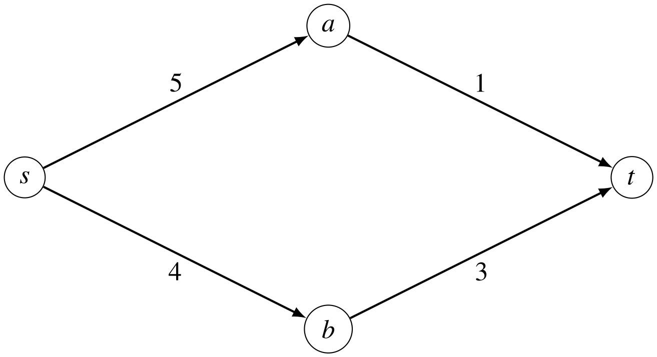

Figure 1. An example routing graph demonstrating the execution of the VCG mechanism

but to *minimize* them as they are cost functions. This results in a different model, and the exact nature of the relation between the two domains is not completely clear at the moment. The MD version of the problem was first studied by Nisan and Ronen in their seminal paper [8] that has arguably been the most influential paper in the area of algorithmic MD. In the scheduling setting we have n machines (players) and m tasks (items). Each machine reports to the mechanism designer the time she would need to process every item, in the form of a time matrix (type profile) $\mathbf{t} = \{t_{ij}\} \subseteq T^{n\times m}$, $T = \mathbb{R}_+$, where t_{ij} is the processing time of machine i for task j. A feasible allocation is an assignment of tasks to players, given (as in the case of additive auctions) by an allocation matrix $\mathbf{a} = \{a_{ij}\} \in \{0,1\}^{n\times m}$, $a_{ij} = 1$ if and only if machine i executes task j. The total valuation of machine i is the sum of the processing times for each individual task assigned (additive valuations) $\sum_{j=1}^{m} a_{ij}t_{ij}$. Each machine's resulting cost is $c_i(\mathbf{t}) = \sum_{j=1}^{m} a_{ij}(\mathbf{t})t_{ij} - p_i(\mathbf{t})$, where $p_i(\mathbf{t})$ represents the payments with which we compensate machine i in order to motivate it to take part in the execution.

Especially with the emergence of the Internet as the predominant computing paradigm, it is natural to assume that these machines will act selfishly and care only about minimizing their own costs $c_i(\mathbf{t})$, and possibly misreport their true processing times to this end. Hence, a game theoretic approach to the classical task allocation problem under the *added truthfulness* constraint is both interesting and necessary. The standard objective is to design truthful mechanisms that minimize the *makespan*

$$\text{Makespan}(\mathbf{t}) = \max_i \sum_{j=1}^{m} a_{ij}(\mathbf{t})t_{ij},$$

that is, the time it would take the slowest machine to finish the processing. Again, we will also require no task to be unprocessed, and thus (4) becomes

$$\sum_{i=1}^{n} a_{ij}(\mathbf{t}) = 1 \quad \text{for all tasks } j \text{ and time matrices } \mathbf{t} \in T^{n\times m}. \tag{7}$$

This is known as the *scheduling problem in parallel unrelated machines* [8,16]. When we consider the *fractional allocations* variant of the scheduling problem [17], we allow $\{a_{ij}\} \in [0,1]^{n\times m}$, while still demanding condition (7). Fractional allocations are essentially randomized mechanisms for the non-fractional version of the problem.

2. Externalities and Challenges

In this chapter we describe the notion of *externalities* in MD settings and present the game theoretic and algorithmic challenges that they introduce. The discussion is inevitably intertwined with challenging the fundamental notion of traditional *truthfulness* itself, and thus, this section essentially forms the motivation for the main results of this chapter, presented in Sec. 3.

2.1. Externalities in Mechanism Design

A fundamental assumption throughout game theory and mechanism design is that all participating players (agents) are fully rational and act selfishly: they have their own well-defined utility function (see e.g. (2)) and they only care about optimizing this function. This can be maximizing satisfaction when they want to buy an item at an auction, or minimizing their cost when they are machines that get to execute tasks (as in the scheduling problem described in Sec. 1.2.3). Usually, this utility function optimization is considered myopic, in the sense that players do not care about variations on the achieved utilities of other players as far as these variations do not affect their own utility levels. For example, in a standard single-item auction setting, if some player does not get the item (thus achieving zero utility), she is indifferent (i.e. her utility does not change) towards how much the winning bidder is going to pay for the item and, even more, she does not care about the bidder's identity.

However, it can be debated whether this really is natural or expected behavior when we think of everyday social interactions. Experiments show that bidders can overbid, possibly risking ending up with negative utilities, just for the joy of winning in case they get the item. Or, on the other hand, if they do not manage to win the item, overbidding will drive prices up in order to harm the other winning agent(s), which is arguably spiteful behavior.

In these examples, where participants in a mechanism behave seemingly irrationally, their happiness is not only a function of their own core utility, but is also affected in complex ways by the other players' utilities. We will call such effects on the modeling of our agents' happiness *externalities*, to emphasize the third party nature of the interaction. Of course, externalities are not only *negative* like in these examples. They can be *positive*, altruistic in nature, e.g. a loving couple taking part in the same auction: one partner may be happy to lose the item if it means that the other one wins it.

Externalities have been heavily studied in economics, not just game theory, and the literature is extensive and diverse. For our purposes, we will approach externalities in the context of the informal definition we gave in the previous paragraphs.

2.2. Impact on Truthfulness

Under the scope of externalities, let us revisit the canonical example of a single-item auction. We know that the second-price paradigm, in particular the Vickrey auction that gives the item to the highest bidding agent but collects as payment the second-highest bid, is optimal for social welfare while also being truthful as no agent has an incentive to lie about her true valuation no matter what the other players report (i.e. truth-telling is a dominant strategy equilibrium). This result allows us not only to maximize the collective happiness of the participants, but also ensure the integrity of the extracted information, here the agents' bids. Furthermore, this is backed up by a very powerful notion of stability for our solution, that of dominant strategies implementation.

However, if we consider spiteful behavior, this does not hold as the second highest bidder, who is going to lose the item anyway, can harm the winning agent by declaring an (untruthful) higher bid that immediately (externally) affects the payment of the other players, and in particular increases the payment for the winning player. Our aim is to study, model, and, if possible, prevent this phenomenon. One simple suggestion would be to encode such possible externalities-behavior into each player's type (some kind of generalized valuation) and then run the powerful VCG mechanisms (see Sec. 1.2.1) on the new extended type-profile space of all agents to get a socially efficient and dominant strategy truthful mechanism. However, there is a fundamental problem with this approach that conflicts all existing tools in the area of mechanism design: the players' utilities now also depend on *other* agents' payments[1]. To be more precise, utilites are no more *quasilinear* functions with respect to payments.

2.3. Challenges

Given the above discussion, the new challenges that are the main motive for this chapter, are the following:

- How to incorporate these externalities properly into the existing standard utility maximization framework of MD? We need a new model for the utility functions, taking into consideration both the core and external utilities of each player.
- Is the standard notion of truthfulness, and in particular that of implementation in dominant strategies, computationally and descriptively appropriate to model this new complex framework of interactions? Can we propose a new notion of empowered truthfulness?
- Utilizing stronger notions of truthfulness, can we design mechanisms that manage to somehow resist externalities that threaten the integrity of traditional truthfulness, the building block of the entire area of MD?
- Is there a general MD paradigm that provides construction of such externality-resistant mechanisms?

3. Strong Truthfulness and Externality Resistance

In this section we start dealing with the challenges discussed in Sec. 2.3. For the remainder of this chapter our exposition is based on the main results of [7]. Further details and missing proofs can be found in that paper.

First, under these new complex underlying interactions among players, we need to have a solid building-block, stronger than the traditional notion of truthfulness which, as we already have seen in Sec. 2.1, even on the simplest example of single-item auctions, can make the powerful VCG second-price auction fail. The intuition is that we would like to introduce a more strict notion of truthfulness, where not only players have no reason to

[1] This is in fact a more appropriate definition, economics-wise, of externalities themselves: utilities are *externally* affected by payments to third parties, over which we have no direct control.

lie but are punished for misreports. We want greater deviations from the true valuations to result in greater reductions to the resulting utilities for the players who deviate.

3.1. The Notion of Strong Truthfulness

Definition 2 (Strong truthfulness)**.** *A mechanism is called c-strongly truthful, for some* $c \in \mathbb{R}$*, if it induces utilities* u_i *such that*

$$u_i(b_i, \mathbf{b}_{-i}) - u_i(\tilde{b}_i, \mathbf{b}_{-i}) \geq \frac{1}{2} c |\tilde{b}_i - b_i|, \tag{8}$$

for every player i, bids b_i*,* $\tilde{b}_i$ *and all bid profiles of the other players* $\mathbf{v}_{-i}$*.*

This notion of strong truthfulness is a generalization of the standard notion of truthfulness, achieved by setting $c = 0$. Under these definitions we can now prove an analytic characterization of strong truthfulness, in the spirit of Thm. 2:

Theorem 3. *A mechanism is c-strongly truthful if and only if the utility functions it induces for every player are all c-strongly convex functions*[2]*.*

3.1.1. Strongly Truthful Auctions

Let us give an example of a strongly truthful auction in the simplest case of a single-buyer, single-item auction. Assume a single player with a valuation for the item drawn from some bounded real interval $[L, H]$. We define the following mechanism, which we will call *linear*.

Definition 3 (Linear mechanism)**.** *The linear mechanism (*LM*) for the single-buyer, single-item auction setting has the allocation*

$$a(b) = \frac{b - L}{H - L} \quad ,$$

where the buyer's bid b for the item ranges over a fixed interval $[L, H]$*.*

It turns out that this mechanism is in fact the strongest possible one we can hope for in this setting.

Theorem 4. *The linear mechanism is* $\frac{1}{H-L}$*-strongly truthful for the single-buyer, single-item auction setting and this is optimal*[3] *among all mechanisms in this setting.*

Notice that Thm. 4 holds only in cases where the valuations domain is a (real) interval $T = [L, H]$. If we want to deal with unbounded domains, e.g. $T = \mathbb{R}_+$, we need to define a more flexible notion of *relative* strong truthfulness (see [7, Sec. 2.2]).

[2] A function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is called *c-strongly convex*, where c is a nonnegative real parameter, if for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$: $f(\mathbf{x}) - f(\mathbf{x}) \geq \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) + \frac{c}{2} \|\mathbf{y} - \mathbf{x}\|^2$, where the standard dot inner product and Euclidean norm are used.

[3] Formally, for every c-strongly truthful mechanism in this setting $c \leq \frac{1}{H-L}$.

3.2. External Utilities Model

In order to model externalities, both positive (altruism) and negative (spite), we are extending our agents' types to include not only their valuations but also some parameters γ that try to quantify the interest/external-effect each agent has upon others. Formally, for every agent $i = 1,2,\dots,n$ we redefine its type t_i to be $t_i = (v_i, \gamma_i)$, where $\gamma_i = (\gamma_{i1}, \gamma_{i2}, \dots, \gamma_{in})$ with $\gamma_{ij} \in \mathbb{R}$, $j = 1,2,\dots,n$ being the *externality parameter* of player i with respect to player j. In fact, parameter γ_{ii} is not needed and we can safely ignore it, setting $\gamma_{ii} = 0$ for all players i, but we still keep it in the formation of γ_i in order to have a vector-consistent notation. The usage of these parameters becomes apparent when we define the utilities in our new externalities model. Intuitively, negative values of γ_{ij} correspond to player i demonstrating spiteful behavior towards player j, positive values correspond to altruistic behavior, and a value of zero corresponds to lack of externalities towards player j (i.e. player i demonstrates standard game-theoretic selfish behavior).

Moving on to defining utilities, let $\mathcal{M} = (\mathbf{a}, \mathbf{p})$ be a mechanism for our new externalities model. Our players have true types $t_i = (v_i, \gamma_i)$ and they submit to the mechanism *bids* (also called *values*) which are (possibly mis-)reports of the valuation component v_i of their true type t_i. Given such a bid profile $\mathbf{b}$, mechanism $\mathcal{M}$ again computes the allocation $\mathbf{a}(\mathbf{b}) = (a_1(\mathbf{b}), a_2(\mathbf{b}), \dots, a_n(\mathbf{b}))$, where $a_i(\mathbf{b}) \in [0,1]$ is the probability that agent i receives the service[4], and the payment vector $\mathbf{p}(\mathbf{b}) = (p_1(\mathbf{b}), p_2(\mathbf{b}), \dots, \mathbf{p}_n(\mathbf{b}))$, where $p_i(\mathbf{b})$ is the payment extracted *from* player i.

In this setting, we define the *base utility* of player i under mechanism $\mathcal{M}$, given the (true) type $t_i = (v_i, \gamma_i)$ and the reported bid vector $\mathbf{b}$, to be the standard utility (1)

$$u_i(\mathbf{b}|t_i) = u_i(\mathbf{b}|v_i) = a_i(\mathbf{b}) \cdot v_i - p_i(\mathbf{b})$$

and then we define her *externality-modified utility*, given also the other players' (true) type profiles $\mathbf{t}_{-i}$, to be

$$\hat{u}_i(\mathbf{b}|\mathbf{t}) = \hat{u}_i(\mathbf{b}|\mathbf{v}, t_i) = u_i(\mathbf{b}|v_i) + \sum_{j \neq i} \gamma_{ij} u_j(\mathbf{b}|t_j) \quad . \tag{9}$$

From now on we will refer to this externality-modified utility simply as utility, as this is going to be the utility notion in our new externalities-included model upon which we will also build our new notions of truthfulness and resistant mechanisms. We observe two important things about these definitions.

First, as expected, the base utility $u_i(\mathbf{b}|\mathbf{t}_i)$ only depends on type $\mathbf{t}_i$ of player i and not on the other players' types $\mathbf{t}_{-i}$ and, in particular, it depends just on the valuation component v_i of $\mathbf{t}_i$ (i.e. the externality parameters $\gamma_i = (\gamma_{i1}, \gamma_{i2}, \dots, \gamma_{in})$ do not play any part in these base utilities). Therefore, we can also use the slightly lighter notation $u_i(\mathbf{b}|v_i)$ to denote the basic utility. Essentially, this is the component of our new utility that corresponds exactly to the standard definition of utilities in the traditional no-externalities setting of MD (see Sec. 1.1).

Second, the externality-modified utility needs to depend on the entire (true) type profile $\mathbf{t}$ and not just the component t_i of i, because the externalities-induced term of equation (9) comprises of a sum that ranges across all other players. Furthermore, unlike

[4] In the single-item auction paradigm, this means agent i gets the item.

for the base utilities, we do not just need the valuations v_i but all parameters γ_{ij} for all $j \neq i$. However, from the parameters profile vector $\mathbf{t}$ we only need the externalities parameters of player i so to make the notation more straightforward, we can write $\hat{u}_i(\mathbf{b}|\mathbf{v},\gamma_i)$ instead of $\hat{u}_i(\mathbf{b}|\mathbf{t})$.

Naturally enough, if $b_i = v_i$, we can denote the base utilities $u_i(\mathbf{b}|\mathbf{t}_i)$ by $u_i(\mathbf{b})$, and if $\mathbf{b} = \mathbf{t}$, we can use $\hat{u}_i(\mathbf{b})$ instead of $\hat{u}_i(\mathbf{b}|\mathbf{t})$ for the externality-modified utilities.

3.3. Externality Resistant Mechanisms

Definition 4 (Externality-resistant VCG). *The externality resistant-VCG mechanism for our MD setting with externalities, denoted by* rVCG(δ) *and parametrized by some* $\delta \in [0,1]$ *is the following protocol:*

- *Ask all n players to report their bids b_i, $i = 1,2,\dots,n$.*
- *With the probability of $\frac{\delta}{n}$ for every player i, single out player i and run the* LM *mechanism (Def. 3).*
- *With complementary probability $1-\delta$, run the standard VCG mechanism.*

Theorem 5. *Consider a MD setting with externalities where players' valuations are drawn from the unit real interval, i.e. $v_i \in [0,1]$ for all $i = 1,2,\dots,m$. Then, for every $\delta,\varepsilon > 0$, if the the externality parameters of our agents satisfy*

$$\max_{i,j} \gamma_{ij} < \frac{\varepsilon\delta}{8(1-\delta)^2 n^3} \ ,$$

the externality-resistant VCG mechanism (Def. 4) induces utilities so that

$$u_i^{\mathrm{rVCG}(\delta)}(\mathbf{b}) \geq (1-\delta)u_i^{\mathrm{VCG}}(\mathbf{v}) - \varepsilon$$

for every player $i = 1,2,\dots,n$ and all undominated *strategy (bid) profiles* $\mathbf{b}$ *of the players.*

The intuition behind this result is that, by randomizing over components or mechanisms that are known to work in simpler settings, the rVCG mechanism manages to achieve (base) utilities that are very close to those of the corresponding VCG mechanism in the no-externalities setting (see Sec. 1.2.1). Of course, this cannot be implemented in dominant strategies (we have seen that truthfulness under such strong solution concepts is doomed to fail) but under a weaker solution concept, that of undominated strategies (see Sec. 3.4).

As an immediate result of this virtual simulation of ideal behavior where only base utilities are taken into consideration, rVCG manages to approximate both optimal social welfare as well as the revenue of the traditional VCG (run on base utilities).

Theorem 6. *In every outcome of* rVCG(δ) *implemented in undominated strategies, the social welfare of* rVCG(δ) *is within an additive error of $n\eta$ from the optimal social welfare and within an additive error of $2n\eta$ of the revenue achieved by the standard* VCG *mechanism (run on base utilies without externalities), where n is the number of players and η is a parameter so that*

$$\eta \leq \frac{4(1-\delta)}{\delta} n^2 \gamma$$

(where $\gamma = \max_{i,j} \gamma_{ij}$).

3.4. Implementation in Undominated Strategies

We briefly describe the solution concept we use in our model and under which our notion of externality resistance is realized, namely *implementation in undominated strategies*. Intuitively, we say that a given property P (here, externality resistance) is implemented in undominated strategies if, for every agent i there is a set of strategies D_i so that playing within D_i is a kind of dominant strategy for every player i. This means that no matter what the other players' strategies are, there is some strategy in D_i that maximizes the utility of player i. In addition, obviously P must be satisfied for all possible input strategies in the product space $\prod_{i=1}^{n} D_i$. For a more formal description we refer to [7, Sec. 1.4] as well as [18] that have utilized this solution concept before. In our specific model, the idea is that as long as the agents stay in strategies that are close enough to truth-telling they are safe. Deviating from the set of strategies D_i will be an absurd choice for agent i, a dominated strategy.

4. Protocol Composability

In this section we explore the bigger picture behind our proposed notion and constructions of Sec. 3. We discuss the underlying schema that achieves resistance towards externalities while still approximating the mechanism designer's objective (e.g. social welfare, Thm. 6).

4.1. Boosting Truthfulness

When we look at Thm. 5, we see that the key property of rVCG (Def. 4) is that the following two values are approximately equal for all agents:

- the utility they end up with in the model *with* externalities after running rVCG, and
- the utility they would have ended up with in an ideal *no*-externalities model after running the traditional VCG mechanism.

In other words, while all agents still bid so as to maximize their new, complex *externality-modified* utilities, they end up with a base utility that is approximately what it would have been if all agents bid so as to maximize their *base* utility. Thus, these externally-resistant mechanisms try to simulate the agents' behavior in an externalities-free utopia and, as a result, they manage to approximate optimal social welfare and revenue.

Above all, what these mechanisms achieve is boosting truthfulness by enforcing incentive-compatibility in this challenging externalities-modified utility model. The key design feature is that of *composability*. If we look at rVCG (Def. 4) we will see that it *randomizes* over *strongly truthful* mechanisms. In particular, it uses the advantage of a strongly truthful mechanism that punishes agents for misbehaving in order to forcefully extract truthful reporting by the agents. It does so by running with some (small) probability such a punishing protocol on a random agent. With the remaining probability we run a standard truthful mechanism that performs optimally with respect to base utilities. In other words, we enrich mechanisms that perform well in the traditional externalities-free model by *composing* them with small, powerful, strongly truthful subroutines.

Such a composable design paradigm, where different mechanisms are combined to boost truthfulness, has been used before in MD settings, e.g. utilizing differential privacy [19], and more subtly in the scoring rules [20,21] and responsive lotteries [22]. However, in [7] the first step is made towards the systematic study of this scheme and the quantification of the performance of the composable mechanisms. Also, this is the first time when this design is used to achieve externality-resistance. Furthermore, our construction has the advantage that it is readily applicable to multidimensional MD settings, such as multi item auctions and scheduling jobs to machines.

4.2. Extensions and Multidimensional Domains

The externality-resistant mechanism rVCG, presented in Sec. 3.3, was applied in a simple, single-dimensional auction setting with only a single item for sale. We want to extend this powerful externality-resistant idea and the composable scheme described in Sec. 4.1 to incorporate more involved *multidimensional* settings with many items, or the scheduling problem.

It turns out that our construction is generic enough to achieve this in a very straightforward way. Consider, for example, the MD scheduling problem of minimizing the makespan of unrelated parallel machines (Sec. 1.2.3). We show how to take advantage of strongly truthful mechanisms to give an unexpected solution to this problem. We will give a mechanism for the problem under the following assumptions:

- The execution times are bounded, in particular we assume that $t_{i,j} \in [L,H]$.
- As in the classical version of the problem, each task must be executed at least once, but in our version it may be executed more than once, even by the same machine [5]. When a machine executes the same task many times, we assume that it pays the same cost t_{ij} for every execution.
- The solution concept for the truthfulness of our mechanism is not dominant strategies but *undominated strategies* (Sec. 3.4).

The mechanism is defined by two parameters $\delta \in [0,1]$ and r.

- The players declare their values $\tilde{t}_{ij}$ that can differ from the real values t_{ij}.
- With the probability $1-\delta$, using the declared values, assign the tasks optimally to the players[6].
- With the remaining probability δ, for every player i and every task, run the truth-extracting LM mechanism (Def. 3), like in the case of the externality-resistant VCG (Def. 4), r times using the declared values $\tilde{t}_{ij}$ from the first step. In fact, we need only to simulate LM once, pretending that the execution time of every task has been scaled up by a factor of r.

[5] The proofs of Nisan and Ronen that give a lower bound of 2 and an upper bound of n for the approximation ratio can be easily extended to this variant of the scheduling problem. The same holds for the lower bound of truthful-in-expectation mechanisms.

[6] Finding or even approximating the optimal allocation with a factor of 1.5 is an NP-hard problem [23], but this is not a concern here, as we focus on the game-theoretic difficulties of the problem. We can replace this part with an approximation algorithm to obtain a polynomial-time approximation mechanism.

Theorem 7. *For every $\delta > 0$ and $\varepsilon > 0$, we can choose the parameter r so that with probability $1-\delta$ the mechanism has the approximation ratio $1+\varepsilon$ and makes no payments; the result holds as long as the players do not play a dominated strategy. This, for example, is achieved for every*

$$r \geq 8n^2 m \frac{H^2}{L^2} \frac{1}{\delta \cdot \varepsilon^2} \quad .$$

Proof. The main idea is that if a machine lies even for one task by more than $\varepsilon_0 = \frac{L}{2n}\varepsilon$, the expected cost of the lie in the truth extraction part of the mechanism will exceed any possible gain. Therefore, the truth-telling strategy dominates any strategy that lies by more than ε_0.

We now proceed with the calculations. If a machine lies about one of its tasks by at least an additive term ε_0, it will pay an expected cost of at least $r\delta\frac{1}{2}\frac{\varepsilon_0^2}{H-L}$. The maximum gain from such a lie is to decrease (with the probability $1-\delta$) its load from mH (the maximum possible makespan) to 0. So the expected gain is at most $(1-\delta)mH \leq mH$, while the loss is at least $r\delta\frac{1}{2}\frac{\varepsilon_0^2}{H-L}$. If we select the parameters so that

$$r\delta\frac{1}{2}\frac{\varepsilon_0^2}{H-L} \geq mH \quad , \tag{10}$$

no machine will have an incentive to lie by more than ε_0, i.e., $|\tilde{t}_{i,j} - t_{i,j}| \leq \varepsilon_0$. But then the makespan computed by the mechanism cannot be more than $m\varepsilon_0$ longer than the optimal makespan: $\text{Makespan}(\tilde{t}) \leq \text{Makespan}(t) + m\varepsilon_0$. We can use the trivial lower bound $\text{Makespan}(t) \geq mL/n$ (or equivalently $n\text{Makespan}(t)/(mL) \geq 1$) to bound the makespan of $\tilde{t}$:

$$\begin{aligned}
\text{Makespan}(\tilde{t}) &\leq \text{Makespan}(t) + m\varepsilon_0 \\
&\leq \text{Makespan}(t) + m\varepsilon_0 \frac{n\text{Makespan}(t)}{mL} \\
&= \left(1 + \frac{n\varepsilon_0}{L}\right)\text{Makespan}(t) \\
&= (1+\varepsilon)\text{Makespan}(t) \quad ,
\end{aligned}$$

where $\varepsilon = \frac{n\varepsilon_0}{L}$. We can make the value of ε as close to 0 as we want by choosing an appropriately high value for r. Constraint (10) shows that $r = \Theta(\delta\varepsilon^{-2})$ is enough. Therefore, with the probability $1-\delta$, the makespan of the declared values is $(1+\varepsilon)$-approximate, for every fixed $\varepsilon > 1$. □

References

[1] John von Neumann and Oskar Morgenstern. *Theory of games and economic behavior*. Princeton University Press, 3rd edition, 1953.

[2] Gillat Kol and Moni Naor. Cryptography and game theory: designing protocols for exchanging information. In *Proceedings of the 5th conference on Theory of cryptography*, TCC'08, pages 320–339, Berlin, Heidelberg, 2008. Springer-Verlag.

[3] Yevgeniy Dodis and Tal Rabin. Cryptography and game theory. In Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay Vazirani, editors, *Algorithmic Game Theory*, chapter 8, pages 181–206. Cambridge University Press, 2007.

[4] Ittai Abraham, Danny Dolev, Rica Gonen, and Joe Halpern. Distributed computing meets game theory: robust mechanisms for rational secret sharing and multiparty computation. In *Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing*, PODC '06, pages 53–62, New York, NY, USA, 2006. ACM.

[5] S. Izmalkov, S. Micali, and M. Lepinski. Rational secure computation and ideal mechanism design. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 585–594. IEEE, 2005.

[6] Y. Dodis, S. Halevi, and T. Rabin. A cryptographic solution to a game theoretic problem. In *Advances in Cryptology—CRYPTO 2000*, pages 112–130. Springer, 2000.

[7] Amos Fiat, Anna R. Karlin, Elias Koutsoupias, and Angelina Vidali. Approaching utopia: Strong truthfulness and externality-resistant mechanisms. In *Innovations in Theoretical Computer Science (ITCS)*, August 2013. http://arxiv.org/abs/1208.3939v1.

[8] N. Nisan and A. Ronen. Algorithmic mechanism design. *Games and Economic Behavior*, 35(1/2):166–196, 2001.

[9] Noam Nisan. Introduction to mechanism design (for computer scientists). In Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay Vazirani, editors, *Algorithmic Game Theory*, chapter 9. Cambridge University Press, 2007.

[10] M.J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1994.

[11] Roger B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, 1981.

[12] Jean-Charles Rochet. The taxation principle and multi-time hamilton-jacobi equations. *Journal of Mathematical Economics*, 14(2):113 – 128, 1985.

[13] William Vickrey. Counterspeculation, auctions and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37, March 1961.

[14] E.H. Clarke. Multipart pricing of public goods. *Public Choice*, 11(1):17–33, 1971.

[15] T. Groves. Incentives in Teams. *Econometrica*, 41(4):617–631, 1973.

[16] George Christodoulou and Elias Koutsoupias. Mechanism design for scheduling. *Bulletin of the EATCS*, 97:40–59, 2009.

[17] George Christodoulou, Elias Koutsoupias, and Annamária Kovács. Mechanism design for fractional scheduling on unrelated machines. *ACM Trans. Algorithms*, 6(2):38:1–38:18, April 2010.

[18] M. Babaioff, R. Lavi, and E. Pavlov. Single-value combinatorial auctions and implementation in undominated strategies. In *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 1054–1063. ACM, 2006.

[19] K. Nissim, R. Smorodinsky, and M. Tennenholtz. Approximately optimal mechanism design via differential privacy. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pages 203–213. ACM, 2012.

[20] G.W. Brier. Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3, 1950.

[21] J. Eric Bickel. Some comparisons among quadratic, spherical, and logarithmic scoring rules. *Decision Analysis*, 4(2):49–65, June 2007.

[22] U. Feige and M. Tennenholtz. Responsive lotteries. *Algorithmic Game Theory*, pages 150–161, 2010.

[23] J.K. Lenstra, D.B. Shmoys, and É. Tardos. Approximation algorithms for scheduling unrelated parallel machines. *Mathematical Programming*, 46(1):259–271, 1990.