Universal Mechanism Design

Sergei Izmalkov, Matt Lepinski and Silvio Micali

June 28, 2005

Simulation of Mediated Games

Our key technical result essentially shows that any finite mediated game G can be transformed to an equivalent (unmediated) ballot-box game Γ.

In a mediated (normal-form) game with incomplete information, each player has his own type and privately communicates a report, an element of his message space, to the mediator, who then evaluates the specified social-choice function on the received reports and publicly announces the outcome *and* any pre-specified information about the reports. (This latter component —empty on one extreme, and consisting of the full reports on the other— formalizes the privacy of a mediated game.) Since the mediator is trusted but the players are not, for completeness, an outcome must be defined in all contingencies, including when a player *aborts*, that is, acts outside his specified boundaries by not sending any report or by sending a report outside his message space —e.g., an insult to the mediator. We thus assume that, when player i aborts, a predetermined outcome Y_i is realized, and that no information about the players' reports is revealed.[1]

A ballot-box game is an ordinary extensive-form game with incomplete information. It starts with a sequence of players' actions which include physical actions such as sealing a message into an envelope and randomizing the order of a set of envelopes via a bingo-like device, *the ballot box*. It ends with a special *swap* operation, in which all the players simultaneously exchange specified sets of sealed envelopes. Each player then learns the outcome by opening all the envelopes in his possession.

By saying that a mediated normal-form game G is equivalent to a ballot-box game Γ, we mean two things. First, that for any equilibrium of G, of any strength (e.g., Bayesian-Nash, dominant strategies, ex post, etc.), there exists an equilibrium of Γ that has the same strength and induces the same outcome function—and vice versa for any equilibrium of Γ. Second, that the privacy of the players' types in G is exactly preserved in Γ: no group of players (whether collaborating or not) may find more information about other players' types in Γ than they may in G.

Implications for Mechanism Design

The problem of finding a mechanism yielding a *social-choice function* μ in an equilibrium (of a given strength) lies at the heart of Mechanism Design.[2] The Revelation principle implies that

[1] Hopefully, such Y_i is an outcome unfavorable to the aborting player. Without loss of any generality, we can think that the communication with the mediator is sequential, so that it is clear who aborts first. For a detailed discussion of this assumption see Section **??**.

[2] A mechanism, or a game form, is a mapping of messages (strategies) into outcomes. A mechanism considered in a given *context* —a description of players (type spaces, utilities) and outcomes— forms a game. A direct mechanism

such a mechanism exists if and only if μ itself is *incentive compatible*—i.e., if truthtelling is an equilibrium of the direct game D_μ in which each player i privately tells his type t_i to a trusted mediator, who then publicly announces $\mu(t_1, \ldots, t_n)$ as an outcome. Direct mechanisms, however, are rather abstract.

Our first contribution to mechanism design consists of proving that, whenever μ is incentive compatible, (1) there always exists a *concrete* mechanism implementing μ, and (2) such a mechanism can be found *automatically*. That is, we provide a universal, computationally efficient procedure that, given any finite, incentive-compatible, social-choice function μ, generates a ballot-box game Γ_μ implementing μ with the same equilibrium strength. (In essence, our procedure simply returns an unmediated, ballot-box game Γ_μ simulating D_μ.)

Whenever μ is dominant-strategy incentive compatible, the direct mechanism induced by μ typically has a unique dominant-strategy equilibrium.[3] A trivial corollary is that then all dominant-strategy equilibria of Γ_μ yield μ.

In general, however, the existence of additional equilibria pose legitimate concern, since some of them may not yield μ. Accordingly, a stronger notion of implementation, *full implementation*, is often adopted, requiring from a mechanism that all equilibria (of a given strength) implement μ. Known full implementation mechanisms, while ingenious, do rely on trusted mediators, and so again are quite abstract.

Our second contribution is making all such existing (finite) mechanisms concrete. Again, we do this by replacing the trusted mediator by a ballot-box game. In a sense, our approach can be construed as a way to simplify the search for a full-implementation mechanism: namely, it suffices for a designer to come up with such a mechanism in an idealized setting, where a trusted mediator is available. Then, our procedure will automatically generate a concrete mechanism.

The Case for Privacy

Privacy of information is clearly a human desideratum, yet it has not received explicit attention in Game Theory. We view people's interest in privacy as stemming from a utilitarian concern about future interactions: information about us could be revealed to our future detriment, or deliberately leaked to our future advantage.

We measure privacy as the amount of information about players' types disclosed during a game: the less information is disclosed, the more privacy is preserved. An important case to consider is that of *maximum possible privacy*. When implementing a given social-choice function μ, such maximum privacy is that corresponding to the mediator of the direct mechanism D_μ, who announces the outcome $y = \mu(\hat{t}_1, \ldots, \hat{t}_n)$ without revealing any additional information about the players' reported types, $\hat{t}_i$.[4] Our simulation result thus implies that maximum privacy can also be obtained by the players alone, that is, by the ballot-box game Γ_μ corresponding to D_μ. In other cases, we might want to reveal more information, but in a precisely controlled way. For instance, if a mechanism designer faces a sequence of related problems, he may want to treat each of the problems separately while

is any mechanism in which the set of messages for each player coincides with his set of types. A social-choice function is a mapping from a vector of players' types into outcomes. Outcome functions that are desirable from economic or social prospective are of particular interest. One example is efficient outcome function—maximizing the sum of player's utilities. For formal definitions see Section **??**.

[3] An exception is when a given player has two or more strategies (types) that he is indifferent to report. Then multiple dominant strategy equilibria are possible, in which outcome functions may differ.

[4] That is, ex ante, each player i knows his own type t_i and the distribution of the types of the others; ex post, he will know y and whatever can be deduced from it about the other players' types, but no more.

controlling how much information flows from one game to the next. Our simulation result allows him to achieve this goal by designing a sequence of mediated normal-form mechanisms that disclose the desired amount of information and then replacing this mediated sequence with a sequence of ballot-box games.

The Implications of Privacy

Today, the only known solutions to many mechanism-design problems require a trusted mediator, but we can trade the amount of trust bestowed on him with the amount of information released. Consider a second-price auction whose mediator announces the winner and the price and (i) nothing more; (ii) all loosing bids; or (iii) all bids. In the third case, all player information is made public, but the mediator is minimally trusted, so that he can be trivially simulated by the players themselves (by just sealing their bids into envelopes, and then publicly opening all of them). We must ask ourselves whether our instinctive preference for more transparent mechanisms might have resulted in choosing outcomes and social-choice functions that tend to reveal more information than strictly necessary. If so, our ability of precisely controlling privacy may actually influence the way in which we choose outcomes and social-choice functions. By guaranteeing that all mediators can be exactly simulated, ballot-box games enable us to dispense with trust-privacy tradeoffs, and to meaningfully consider alternative choices. For instance, auctions (i) and (ii) can be implemented by the players themselves with equal transparency. As for another, more important example, ballot-box mechanisms enable us to implement social-choice functions with "private outcomes." In particular, assuming that actual transactions can be kept private, we can implement second-price auctions where the losers learn only that they have lost, the winner learns that he has won and at what price, and the seller learns only the winner and the price.

Implications for Secure Computation

Our result succeed in combining incentive compatibility with *Secure Function Evaluation* (SFE). A general SFE protocol enables a group of n players — each possessing his own private input x_i — to talk back and forth so as to evaluate on these inputs any finite (n-valued) function f as if in the *ideal evaluation*, where each player i hands x_i to a trusted party, who then evaluates $f(x_1, \ldots, x_n) = (y_1, \ldots, y_n)$ and returns y_i to i. Put forward by Goldreich, Micali, and Wigderson (1987) —improving on 2-party results of Yao (1986)— this general notion has been extensively studied and achieved with different assumptions, communication channels, and security notions. Despite so much attention, however, no prior SFE protocol preserves f's incentives (unless f is incentive compatible in dominant strategies). By this we mean that, though rational players choose to report their original inputs in an ideal evaluation of f, they will *not* choose to run the SFE protocol on the same inputs.[5]

By simulating any mediated game exactly, we actually obtain a novel type of SFE protocol, which enjoys extraordinary security properties. In addition to not relying on any complexity assumptions and enjoying all previously achieved security properties, ours is the first SFE protocol that preserves the incentive compatibility of any function f.

[5] As we shall explain, this is so because prior protocols: (1) rely on at least some players being *honest* —i.e., blindly following his prescribed strategy— rather than being fully rational; and (2) cannot preclude *signalling,* —i.e., cannot preclude players from volunteering arbitrary information about their own types to other players.

References

Goldreich, O., S. Micali, and A. Wigderson (1987): "How to play any mental game," in *STOC '87*, pp. 218–229. ACM.

Yao, A. (1982): "Protocols for Secure Computations," in *Proc. of FOCS '82*. IEEE.