Bridging Game Theory and Cryptography: Recent Results and Future Directions

Jonathan Katz*

Department of Computer Science
University of Maryland
jkatz@cs.umd.edu

Abstract. Motivated by the desire to develop more realistic models of, and protocols for, interactions between mutually distrusting parties, there has recently been significant interest in combining the approaches and techniques of game theory with those of cryptographic protocol design. Broadly speaking, two directions are currently being pursued:

Applying cryptography to game theory: Certain game-theoretic equilibria are achievable if a trusted *mediator* is available. The question here is: *to what extent can this mediator be replaced by a distributed cryptographic protocol run by the parties themselves?*

Applying game-theory to cryptography: Traditional cryptographic models assume some honest parties who faithfully follow the protocol, and some arbitrarily malicious players against whom the honest players must be protected. Game-theoretic models propose instead that all players are simply *self-interested* (i.e., rational), and the question then is: *how can we model and design meaningful protocols for such a setting?*

In addition to surveying known results in each of the above areas, I suggest some new definitions along with avenues for future research.

1 Introduction

The fields of *game theory* and *cryptographic protocol design* are both concerned with the study of "interactions" among mutually distrusting parties. These two subjects have, historically, developed almost entirely independently within different research communities and, indeed, they tend to have a very different flavor. Recently, however, motivated by the desire to develop more realistic models of (and protocols for) such interactions, there has been significant interest in combining the techniques and approaches of both fields.

Current research at the intersection of game theory and cryptography can be classified into two broad categories: applying cryptographic protocols to game-theoretic problems, and applying game-theoretic models and definitions to the general area of cryptographic protocol design. In a bit more detail:

* Research supported in part by the U.S. Army Research Laboratory, NSF CAREER award #0447075, and US-Israel Binational Science Foundation grant #2004240.

– Certain game-theoretic equilibria are possible if parties rely on the existence of an external trusted party called a *mediator*. (All the relevant definitions are given in Section 2.) This naturally motivates a cryptographer[1] to ask: *can the trusted mediator be replaced by a protocol that is run by the parties themselves*? Research aimed at understanding the conditions under which the answer is positive, and developing appropriate protocols in such cases, is described in Section 3.
– Traditionally, cryptographic protocols are designed under the assumption that some parties are *honest* and faithfully follow the protocol, while some parties are *malicious* and behave in an arbitrary fashion. The game-theoretic perspective, however, is that all parties are simply *rational* and behave in their own best interests. This viewpoint is incomparable to the cryptographic one: although no one can be trusted to follow the protocol (unless it is in their own best interests), the protocol need not prevent "irrational" behavior. The general question here is: *what models and protocols are appropriate for this setting?* This work is discussed in Section 4.

This paper surveys recent work in both the directions listed above, with a cryptographic audience in mind. This survey focuses more on the problems being addressed than on the solutions that have been proposed, and will thus emphasize definitions rather than concrete results. I also propose new definitional approaches to some of the problems under discussion, and have made a particular effort to highlight promising directions for future research.

Dodis and Rabin have recently written an excellent survey [16] that covers very similar ground as the present work. The present survey is perhaps a bit more technical, and somewhat more opinionated. Surveys more tangentially related to the topics considered here include those by Linial [33] and Halpern [25].

It is fascinating to observe that the recent growth of interest in blending game theory and cryptography has paralleled a surge of attention focused on game theory by computer scientists in general, most notably (for the purposes of this work) in the fields of computational complexity (see, e.g., [38, Chap. 2]), networking and distributed algorithms (see, e.g., [38, Chap. 14]), network security (see, e.g., [10] and [38, Chaps. 23, 27]), information security economics [38, Chap. 25], and more. These are all well beyond the scope of the present work.

Note: Due to space limitations, this survey has been shortened somewhat. A full version will be posted and maintained at `http://eprint.iacr.org`. Comments and corrections are very much appreciated.

2 A Crash Course in Game Theory

This section reviews some central game-theoretic concepts. I have tried to simplify things when, in my view, nothing of essence is lost (vis-a-vis the results presented here). For extensive further details, the reader is referred to [39, 19].

[1] Although, interestingly, the question was first asked in the economics community.

We begin by introducing the notion of *normal form games.* A n-player game $\Gamma = (\{A_i\}_{i=1}^n, \{u_i\}_{i=1}^n)$, presented in normal form, is determined by specifying, for each player P_i, a set of possible *actions* A_i and a *utility function* $u_i \colon A_1 \times \cdots \times A_n \mapsto \mathbb{R}$. Letting $A \stackrel{\text{def}}{=} A_1 \times \cdots \times A_n$, we refer to a tuple of actions $\boldsymbol{a} = (a_1, \ldots, a_n) \in A$ as an *outcome.* The utility function u_i of party P_i expresses this player's preferences over outcomes: P_i prefers outcome $\boldsymbol{a}$ to outcome $\boldsymbol{a'}$ iff $u_i(\boldsymbol{a}) > u_i(\boldsymbol{a'})$. (We also say that P_i *weakly* prefers $\boldsymbol{a}$ to $\boldsymbol{a'}$ if $u_i(\boldsymbol{a}) \geq u_i(\boldsymbol{a'})$.) We assume that the $\{A_i\}, \{u_i\}$ are common knowledge among the players, although the assumption of known utilities seems rather strong and it is preferable to avoid it (or assume only limited knowledge).

The game is played by having each party P_i select an action $a_i \in A_i$, and then having all parties play their actions *simultaneously.* The "payoff" to P_i is given by $u_i(a_1, \ldots, a_n)$ and, as noted above, P_i is trying to maximize this value.

Two-player games (for reasonably sized A_1, A_2) can be represented conveniently in matrix form by labeling the rows (resp., columns) of the matrix with the actions in A_1 (resp., A_2). The entry in the cell at row $a_1 \in A_1$ and column $a_2 \in A_2$ contains a tuple (u_1, u_2) indicating the payoffs to P_1 and P_2, respectively, given the outcome $\boldsymbol{a} = (a_1, a_2)$. For example, the following represents a game where $A_1 = \{C, D\}$, $A_2 = \{C', D'\}$, and, e.g., $u_1(C, D') = 1$ and $u_2(C, D') = 3$:

	C'	D'
C	$(2,2)$	$(1,3)$
D	$(3,1)$	$(0,0)$

Table 1. A two-player game.

Types, and games of incomplete information. The above definition corresponds to so-called games of *perfect* (or *complete*) information. One can also consider extensions that model different features of "real-world" interactions, such as inputs provided to the parties at the beginning of the game whose values affect players' utilities. (In the game theory literature these inputs are said to determine the *type* of each party.) We now provide a simplified definition incorporating this situation; see [39, 19] for the general case.

Let $\Gamma = (\{A_i\}, \{u_i\})$ be as above, where the $\{u_i\}$ are now functions from $(\{0,1\}^*)^n \times A$ to the reals. Let $\mathcal{D}$ be a distribution over vectors $(t_1, \ldots, t_n)$, where each t_i is a binary string. A game is now played as follows: first, $(t_1, \ldots, t_n)$ is sampled according to $\mathcal{D}$, and P_i is given t_i. Next, each player P_i plays an action $a_i \in A_i$ as before; once again, these are all assumed to played simultaneously. Then, each player P_i receives payoff $u_i(t_1, \ldots, t_n, a_1, \ldots, a_n)$.

2.1 Nash Equilibria

If parties play a game (of perfect information), what can we expect to happen? Say P_1 knows the actions $a_2, \ldots, a_n$ that the other parties are going to take. Then it will choose the action $a_1 \in A_1$ that maximizes $u_1(a_1, \ldots, a_n)$; we call this a_1 a *best response* of P_1 to the actions of the other players. (A best response need not be unique.) Given this action chosen by the first player, P_2 will then choose a best response $a'_2 \in A_2$, and so on. We see that a tuple $\boldsymbol{a} = (a_1, \ldots, a_n)$ is "self-enforcing" only if each a_i represents P_i's best response to $\boldsymbol{a}_{-i} \stackrel{\text{def}}{=} (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n)$. A tuple with this property is called a *Nash equilibrium*, and this serves as the starting point for all further analysis of the game. Formally, if we let $(a'_i, \boldsymbol{a}_i)$ denote $(a_1, \ldots, a_{i-1}, a'_i, a_{i+1}, \ldots, a_n)$, we have:

Definition 1. *Let $\Gamma = (\{A_i\}_{i=1}^n, \{u_i\}_{i=1}^n)$ be a game presented in normal form, and let $A = A_1 \times \cdots \times A_n$. A tuple $\boldsymbol{a} = (a_1, \ldots, a_n) \in A$ is a* (pure-strategy) Nash equilibrium *if for all i and any $a'_i \in A_i$ it holds that $u_i(a'_i, \boldsymbol{a}_{-i}) \leq u_i(\boldsymbol{a})$.*

Another way of expressing this is to say that $a_i \in A_i$ *weakly dominates* $a'_i \in A_i$ *relative to* $\boldsymbol{a}_{-i}$ if $u_i(a_i, \boldsymbol{a}_{-i}) \geq u_i(a'_i, \boldsymbol{a}_{-i})$. Then $\boldsymbol{a}$ is a Nash equilibrium if, for all i, the action a_i weakly dominates all actions in A_i relative to $\boldsymbol{a}_{-i}$.

In the example of Table 1, (C, D') is a pure-strategy Nash equilibrium: given that P_1 plays C, the second player prefers to play D'; given that P_2 plays D', the first player prefers to play C. A second Nash equilibrium is given by (D, C').

In the above definition of a pure-strategy Nash equilibrium, the "strategy" of P_i was to deterministically play a_i (hence the name *pure* strategy). If we limit players to such strategies, a Nash equilibrium may not exist in a given game. To remedy this, we allow players to follow *randomized* strategies as well. Specifically, if σ_i is a probability distribution over A_i then we also let σ_i represent the strategy in which P_i samples $a_i \in A_i$ according to σ_i and then plays this action. (We recover deterministic strategies by letting σ_i assign probability 1 to some action.) Given a strategy vector $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$, we overload notation and let $u_i(\boldsymbol{\sigma})$ denote the *expected* utility of P_i given that all parties play according to $\boldsymbol{\sigma}$. (We remark that although this is the standard way to assign utilities to distributions over outcomes, doing so makes the generally unrealistic assumption that players are *risk neutral* in that they care only about their expected utility.) The strategy σ_i is a *best response* to $\boldsymbol{\sigma}_{-i}$ if it maximizes $u_i(\sigma_i, \boldsymbol{\sigma}_{-i})$. Then:

Definition 2. *Let $\Gamma = (\{A_i\}_{i=1}^n, \{u_i\}_{i=1}^n)$ be as above, and let σ_i be a distribution over A_i. Then $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$ is a* (mixed-strategy) Nash equilibrium *if for all i and any distribution σ'_i over A_i it holds that $u_i(\sigma'_i, \boldsymbol{\sigma}_{-i}) \leq u_i(\boldsymbol{\sigma})$.*

One can verify that in the two-party game of Table 1, the strategy vector in which P_1 plays C with probability 1/2, and in which P_2 plays C' with probability 1/2 is a (mixed-strategy) Nash equilibrium.

The celebrated theorem of Nash [37] is that any game of perfect information where the $\{A_i\}$ are finite has a (mixed-strategy) Nash equilibrium. The finiteness assumption is necessary, as there are examples of two-player games with countably-infinite action sets where no mixed-strategy Nash equilibrium exists.

Nash equilibria for games of incomplete information can be defined in the natural way based on the above. Here the strategy of player P_i corresponds to a *function* mapping its received input t_i to an action $a_i \in A_i$; pure strategies correspond to deterministic functions. Note that here we must take into account parties' *expected* utilities even when considering pure-strategy Nash equilibria, since the utility of P_i may depend on the types of the other players, and these are unknown at the time P_i chooses its action.

2.2 Other Equilibrium Concepts

Nash equilibria are considered by many to be *the* fundamental equilibrium notion for games. Nevertheless, it is of interest to explore various refinements and strengthenings of this concept.

Dominated strategies and iterated deletion. Given a game $\Gamma = (\{A_i\}, \{u_i\})$, we say that action $a_i \in A_i$ is *strictly dominated* with respect to A_{-i} if there exists a randomized strategy $\sigma_i \in \Delta(A_i)$ such that $u_i(\sigma_i, \boldsymbol{a}_{-i}) > u_i(a_i, \boldsymbol{a}_{-i})$ for all $\boldsymbol{a}_{-i} \in A_{-i}$ (where $A_{-i} \stackrel{\text{def}}{=} \times_{j \neq i} A_j$). I.e., a_i is strictly dominated if P_i can always improve its situation by *not* playing a_i. An action $a_i \in A_i$ is *weakly dominated* with respect to A_{-i} if there exists a randomized strategy $\sigma_i \in \Delta(A_i)$ such that (1) $u_i(\sigma_i, \boldsymbol{a}_{-i}) \geq u_i(a_i, \boldsymbol{a}_{-i})$ for all $\boldsymbol{a}_{-i} \in A_{-i}$, and (2) there exists $\boldsymbol{a}_{-i} \in A_{-i}$ such that $u_i(\sigma_i, \boldsymbol{a}_{-i}) > u_i(a_i, \boldsymbol{a}_{-i})$. I.e., P_i can never improve its situation by playing a_i, and can sometimes improve its situation by not playing a_i.

It seems that a rational player will never choose a strictly dominated action. In fact, it is not hard to show that in any Nash equilibrium, no player assigns positive probability to any strictly dominated action. Arguably, a rational player should also never choose a weakly dominated action (although the argument in this case is less clear). If we accept this assumption, then a Nash equilibrium in which some party plays a weakly dominated action with positive probability is not expected to occur in practice. For example, consider the following game:

	C'	D'
C	$(10, 10)$	$(1, 1)$
D	$(10, 0)$	$(2, 2)$

(C, C') is a Nash equilibrium. However, action C of player P_1 is weakly dominated by action D. Thus, we may expect that P_1 plays D — but this forces us to the Nash equilibrium (D, D'). Note that both players now end up doing worse! Intuitively, both players prefer the Nash equilibrium (C, C'), but this is not "stable" in a sense we will define below.

Say we are given a game Γ^0, and we have eliminated the weakly dominated actions of each player from consideration. This leaves us with "effective" action sets $\{A_i^1\}$ for each player. We may now iterate the process, and remove any actions that are weakly dominated in the "reduced game" $\Gamma^1 = \big(\{A_i^1\}, \{u_i\}\big)$, etc. This leads to the following definition.

Definition 3. *Given $\Gamma = (\{A_i\}, \{u_i\})$ and $\hat{A} \subseteq A$, let $\mathsf{DOM}_i(\hat{A})$ denote the set of strategies in $\hat{A}_i$ that are weakly dominated with respect to $\hat{A}_{-i}$. For $k \geq 1$, set $A_i^k \stackrel{\text{def}}{=} A_i^{k-1} \setminus \mathsf{DOM}_i(A^{k-1})$. Set $A_i^\infty \stackrel{\text{def}}{=} \cap_k A_i^k$. A Nash equilibrium $\boldsymbol{\sigma}$ of Γ* survives iterated deletion of weakly dominated strategies *if $\sigma_i \in \Delta(A_i^\infty)$ for all i.*

Stability with respect to trembles. Another means to distinguish among a set of Nash equilibria is to ask how stable each such equilibrium is to "mistakes" (or *trembles*) of the other players. Such mistakes might correspond to a real mistake on the part of some player (e.g., a player chooses an irrational strategy by accident), some "out-of-band" event (e..g, a network failure), or the fact that a player's utility is slightly different than originally thought.

To define stability with respect to trembles, we must first define a metric d on the strategy space $\Delta(A)$ of the players. Assuming A is finite, a natural candidate is statistical difference and we assume this in the definition that follows. Various notions of stability with respect to trembles have been considered in the game theory literature, although some of them seem problematic in a cryptographic setting. The following seems best for our context:

Definition 4. *Let $\Gamma = (\{A_i\}, \{u_i\})$, and let $\boldsymbol{\sigma}$ be a Nash equilibrium in Γ. Then $\boldsymbol{\sigma}$ is* stable with respect to trembles *if there exists an $\epsilon > 0$ such that for all i and every $\boldsymbol{\sigma}'_{-i} \in \Delta(A_{-i})$ with $d(\boldsymbol{\sigma}_{-i}, \boldsymbol{\sigma}'_{-i}) < \epsilon$, the strategy σ_i is a best response to $\boldsymbol{\sigma}'_{-i}$. I.e., for every $\sigma'_i \in \Delta(A_i)$ it holds that $u_i(\sigma'_i, \boldsymbol{\sigma}'_{-i}) \leq u_i(\sigma_i, \boldsymbol{\sigma}'_{-i})$.*

That is, even if P_i believes there is some small probability that the other players will make a mistake (and not play according to $\boldsymbol{\sigma}_{-i}$), it is still in P_i's best interests to play according to σ_i.

As an example, consider the following two-player game:

	A'	B'	C'
A	$(10, 2)$	$(1, 0)$	$(0, 1)$
B	$(10, 0)$	$(0, 0)$	$(100, 100)$

(A, A') is a Nash equilibrium, but it is not stable with respect to trembles: if P_1 believes that P_2 might play C' with any positive probability ϵ (but still plays B' with probability 0), then P_1 will prefer to play B rather than A. On the other hand, (C, C') is a Nash equilibrium that *is* stable with respect to trembles: for small enough $\epsilon > 0$, even if P_1 believes that P_2 might play something other than C' with probability ϵ, it is still in P_1's best interest to play C.

I am not aware of any results stating conditions under which stable equilibria are guaranteed to exist.

Coalitions. Thus far, we have only been considering single-player deviations, i.e., whether it is in any single player's best interests to deviate from some prescribed strategy. Cryptographers generally prefer to think in terms of coalitions of players acting together. In general, a Nash equilibrium provides no "protection" against such coalitions.

What does it mean for a coalition $\mathcal{C}$ to prefer one outcome to another? There are at least four natural possibilities:

– $\mathcal{C}$ prefers $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma}'$ only if *every* player in $\mathcal{C}$ weakly prefers $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma}'$, and some player in $\mathcal{C}$ strictly prefers $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma}'$.

- $\mathcal{C}$ prefers $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma}'$ only if the sum of the utilities of the parties in $\mathcal{C}$ improves; i.e., if $\sum_{i \in \mathcal{C}} u_i(\boldsymbol{\sigma}) > \sum_{i \in \mathcal{C}} u_i(\boldsymbol{\sigma}')$. (Note that for this to make sense, we must assume that the utility functions of all players in $\mathcal{C}$ are measured in the same units.) This definition can be viewed as capturing the ability of players in $\mathcal{C}$ to make "side payments" to each other before or after the game.
- $\mathcal{C}$ prefers $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma}'$ if *any* player in $\mathcal{C}$ prefers $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma}'$, i.e., if $u_i(\boldsymbol{\sigma}) > u_i(\boldsymbol{\sigma}')$ for some $i \in \mathcal{C}$. The definition makes sense if we think of one adversarial party corrupting other parties and taking complete control over their actions. Note also that preference of $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma}'$ with respect to this definition implies preference with respect to the previous two definitions.
- Another possibility is to simply assume utility functions $u_{\mathcal{C}}$ for each possible coalition $\mathcal{C}$, and then define preference in the obvious way. This is the most general approach (it subsumes the previous three), but requires additional assumptions about players' utilities.

We adopt the third definition here.

Given a set $\mathcal{C} = \{i_1, \ldots, i_t\} \subset [n]$ and a vector $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$, we let $A_{\mathcal{C}} \stackrel{\text{def}}{=} \times_{i \in \mathcal{C}} A_i$, $\sigma_{\mathcal{C}} \stackrel{\text{def}}{=} (\sigma_{i_1}, \ldots, \sigma_{i_t})$, and $\sigma_{-\mathcal{C}} \stackrel{\text{def}}{=} \sigma_{[n] \setminus \mathcal{C}}$. Then:

Definition 5. *Let $\Gamma = (\{A_i\}_{i=1}^n, \{u_i\}_{i=1}^n)$. Then for $1 \leq t < n$ the strategy vector $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$ is a t-resilient equilibrium if for all $\mathcal{C} \subset [n]$ with $|\mathcal{C}| \leq t$, all $i \in \mathcal{C}$, and any $\boldsymbol{\sigma}'_{\mathcal{C}} \in \Delta(A_{\mathcal{C}})$, it holds that $u_i(\boldsymbol{\sigma}'_{\mathcal{C}}, \boldsymbol{\sigma}_{-\mathcal{C}}) \leq u_i(\boldsymbol{\sigma})$.*

That is, for every coalition $\mathcal{C}$ of size at most t, no member of the coalition improves its situation no matter how the members of $\mathcal{C}$ coordinate their actions.

Observe that a 1-resilient equilibrium is a Nash equilibrium. Extending other equilibrium concepts to the case of coalitions seems not to have been explored significantly.

Mixed models. It is standard in game theory to assume that all players are rational. Recent work [1, 35, 2] has explored models where most parties are rational, but some players are *malicious* and behave arbitrarily. Treating players as malicious can be viewed (to some extent) as treating their utilities as completely unknown. It is also possible to assume that some players honestly follow the proscribed protocol — perhaps out of altruism or laziness — rather than seeking to improve their utility (although it should be in these players' interests to run the protocol altogether). These are interesting directions that are not discussed any further in this survey.

2.3 Correlated Equilibria

Correlated equilibria [3] offer another solution concept with some advantages relative to Nash equilibria. In some games, there may exist a correlated equilibrium that, for every party P_i, gives a better payoff to P_i than any Nash equilibrium (see [36] for an example). More generally, correlated equilibria have payoffs outside the convex hull of all Nash equilibria, and therefore give more options to the players. Finally, correlated equilibria of any game can be computed in polynomial time, something not believed to be the case for Nash equilibria.

Given a game $\Gamma = (\{A_i\}, \{u_i\})$, we define a *mediated* version of Γ which relies on a trusted, external party M called the *mediator*. The game is now played in two stages: first, the mediator chooses a vector of actions $\boldsymbol{a} \in A$ according to some known distribution $\mathcal{M}$, and then hands the *recommendation* a_i to player P_i. The players then play Γ as before by choosing any action in their respective action sets. Players are "supposed" to follow the recommendation of the mediator, and a *correlated equilibrium* is one in which it is in each player's best interests to do so. To formally define this notion, let $u_i(a'_i, \boldsymbol{a}_{-i} \mid a_i)$ denote the expected utility of P_i, given that it plays action a'_i after having received recommendation a_i and all other parties play their recommended actions $\boldsymbol{a}_{-i}$. (The expectation here is over $\boldsymbol{a}$ sampled according to $\mathcal{M}$.)

Definition 6. *Let $\Gamma = (\{A_i\}, \{u_i\})$. A distribution $\mathcal{M} \in \Delta(A)$ is a* correlated equilibrium *if for all $\boldsymbol{a} = (a_1, \ldots, a_n)$ in the support of $\mathcal{M}$, all i, and all $a'_i \in A_i$, it holds that $u_i(a'_i, \boldsymbol{a}_{-i} \mid a_i) \leq u_i(\boldsymbol{a} \mid a_i)$.*

Any Nash equilibrium is a correlated equilibrium, but Nash equilibria correspond to the special case where $\mathcal{M}$ is a *product distribution* over the A_i.

Let $u_i(\mathcal{M})$ denote the expected utility of P_i when all parties follow their actions as recommended by $\mathcal{M}$. A definition equivalent to the previous one, but better suited for extensions to coalitions as well as the computational setting, is:

Definition 7. *Let $\Gamma = (\{A_i\}, \{u_i\})$. A distribution $\mathcal{M} \in \Delta(A)$ is a* correlated equilibrium *if for all i and any $f_i : A_i \to A_i$ it holds that*

$$u_i(f_i(a_i), \boldsymbol{a}_{-i}) \leq u_i(\mathcal{M}),$$

where $\boldsymbol{a}$ is sampled according to $\mathcal{M}$.

As an example of a game with a correlated equilibrium that is not a Nash equilibrium, consider the two-party game of Table 1 and the distribution that assigns probability 1/3 to each of (C, D'), (D, C'), and (C, C'). One can check that neither party has any incentive to deviate from their recommended action, and each player has expected utility 2 (an improvement on the mixed-strategy Nash equilibrium described in Section 2.1).

In games of incomplete information (as we have defined them in Section 2), a mediated game is played as follows: first, a vector $(t_1, \ldots, t_n)$ is sampled according to a distribution $\mathcal{D}$, and t_i is given to P_i. Then, each party P_i sends some t'_i to the mediator. Based on the vector $\boldsymbol{t'} = (t'_1, \ldots, t'_n)$ received, the mediator samples a vector $\boldsymbol{a} \in A$ according to a distribution $\mathcal{M}(\boldsymbol{t'})$, and recommends action a_i to player P_i. The parties then play as before, choosing whether or not to follow the mediator's recommendation. Correlated equilibria in this situation are defined as the natural extension of the above, through we stress that a player's strategy now determines *both* what value t'_i it sends to the mediator (as a function of the received input t_i) as well as what action it plays in the game. A correlated equilibrium is said to be *truthful* if it is in each party's best interest to send $t'_i = t_i$ to the mediator. The *revelation principle* characterizes when truthful correlated equilibria exist.

Correlated equilibria in the presence of coalitions. The basic approach used to handle coalitions in Definition 5 can be extended in the natural way to the case of correlated equilibria. Two variants of the definition are obtained, however, depending on the details of how the mediated game is played. If *ex ante* collusion is allowed, the parties in a coalition $\mathcal{C}$ may coordinate their strategies in advance, but are assumed unable to communicate after the mediator provides them with their recommended actions. If *ex post* collusion is allowed, the parties in $\mathcal{C}$ can communicate even after receiving their recommendations from the mediator.

Definition 8. *Let $\Gamma = (\{A_i\}, \{u_i\})$ be an n-party game, and let $1 \leq k < n$. A distribution $\mathcal{M} \in \Delta(A)$ is an* ex ante t-resilient correlated equilibrium *if for all $\mathcal{C} \subset [n]$ with $|\mathcal{C}| \leq k$, any functions $\{f_i \colon A_i \to A_i\}_{i \in \mathcal{C}}$, and all $i \in \mathcal{C}$ it holds that $u_i\left(\{f_i(a_i)\}_{i \in \mathcal{C}}, \boldsymbol{a}_{-\mathcal{C}}\right) \leq u_i(\mathcal{M})$, where $\boldsymbol{a}$ is sampled according to $\mathcal{M}$.*

$\mathcal{M}$ is an ex post t-resilient correlated equilibrium *if for all $\mathcal{C}$ as above, any function $f_{\mathcal{C}} : A_{\mathcal{C}} \to A_{\mathcal{C}}$, and any $i \in \mathcal{C}$ it holds that $u_i(f_{\mathcal{C}}(\boldsymbol{a}_{\mathcal{C}}), \boldsymbol{a}_{-\mathcal{C}}) \leq u_i(\mathcal{M})$, where $\boldsymbol{a}$ is sampled according to $\mathcal{M}$.*

2.4 Extensive Form Games

Extensive form games remove the assumption that players act simultaneously. Such games are best described as occurring in a sequence of rounds, where in any given round the game might specify that all parties play simultaneously (as in a normal form game) or that some subset of designated parties plays. Play of the game thus defines a *history* of the actions taken by the players thus far, and a player P_i's strategy σ_i now specifies, for each round in which it is P_i's turn to move, a (randomized) function mapping possible histories to actions. Players' utilities are now functions of terminal histories (i.e., histories that occur at the end of the game), rather than functions of the strategy vector of the players. We rely on the above intuitive description rather than present a formal definition.

We provide a simple example of an extensive form game, which also demonstrates how introducing alternation can affect the outcome of a game. Consider a seller P_1 and a buyer P_2, where P_1 can either sell high (H) or low (L), and P_2 can choose either to buy (B) or not (N). Payoffs are given by the matrix on the left, but we will assume that the seller announces its action first. This gives an extensive form game in which the buyer can follow any of four (pure) strategies; we let XY denote the strategy where P_2 chooses X if the seller chooses H, and P_2 chooses Y if the seller chooses L. This extensive form game is represented in normal form in the matrix on the right.

	B	N
H	$(10, 1)$	$(0, 0)$
L	$(5, 6)$	$(0, 0)$

	BB	BN	NB	NN
H	$(10, 1)$	$(10, 1)$	$(0, 0)$	$(0, 0)$
L	$(5, 6)$	$(0, 0)$	$(5, 6)$	$(0, 0)$

Table 2. An extensive form game presented in normal form.

Looking at the game on the right, we see that (L, NB) is a Nash equilibrium. The strategy being followed by P_2 is to refuse to buy if the buyer charges the higher price, and if the seller knows that P_2 will follow this strategy then it is in the seller's best interest to charge a low price. In contrast, the game on the left (where parties move simultaneously) has the unique Nash equilibrium (H, B).

Something odd about the Nash equilibrium (L, NB) of the extensive form game is that P_2 is, in essence, threatening to play *irrationally* if P_1 plays H (since, conditioned on P_1 playing H, the buyer is better off playing B than N). Another way to say this is that P_2 plays rationally given any realizable history (where history h is *realizable with respect to* $\boldsymbol{\sigma}$ if this history occurs with positive probability when all parties play according to $\boldsymbol{\sigma}$), but P_2 threatens to play irrationally at some non-realizable history. In the following section, we will discuss a refinement of Nash equilibria that eliminates such "empty threat" strategies.

2.5 Equilibrium Concepts in Extensive Form Games

Any game in extensive form can be viewed as a normal form game by letting the set of allowable actions correspond to the players' strategies. Thus, all the equilibrium concepts we have discussed previously can be applied to extensive form games as well. However, it is often more natural to view certain games in extensive form, and thinking of games in this way motivates new equilibrium concepts. In particular, a question that arises with regard to extensive form games is whether we need to "pay attention" to players' strategies at non-realizable histories. In some cases paying attention to such strategies makes intuitive sense, while in other cases the situation is less clear.

Subgame perfect equilibria. As noticed in the previous section, certain strategy vectors may be Nash equilibria but contain "empty threats" by one or more of the players. Subgame perfect Nash equilibria eliminate this possibility. To define this concept, we introduce (informally) the notion of the *reduced game* Γ^h of an extensive form game Γ. Basically, Γ^h corresponds to Γ where some initial history h is fixed; we may view Γ^h as the continuation of Γ conditioned on the fact that history h has been observed thus far. A strategy σ_i in Γ naturally induces a strategy σ_i^h in Γ^h by setting $\sigma_i^h(h') = \sigma_i(h \| h')$.

Definition 9. *Let Γ be an extensive form game, and let $\boldsymbol{\sigma}$ be a Nash equilibrium in Γ. Then $\boldsymbol{\sigma}$ is* subgame perfect *if for all possible histories h of Γ, the strategy vector $\boldsymbol{\sigma}^h$ is a Nash equilibrium of the reduced game Γ^h.*

Recall that a history h is realizable (with respect to $\boldsymbol{\sigma}$) if it occurs with positive probability when all parties follow $\boldsymbol{\sigma}$. If the definition above only quantified over realizable histories, then every Nash equilibrium would satisfy the definition.

In the game of Table 2 the Nash equilibrium (L, NB) is not subgame perfect because, conditioned on the (non-realizable) history in which P_1 plays H, player P_2 prefers to play B instead of N. Equilibrium (H, BB) is subgame perfect.

Stability with respect to trembles. There are two possible ways to extend the definition of stability with respect to trembles to extensive form games, depending on whether or not subgame perfection is also required. The following definition does *not* take subgame perfection into account. Say two strategies σ_i, σ'_i of P_i *yield equivalent play with respect to* $\boldsymbol{\sigma}$ if for every history h realizable with respect to $\boldsymbol{\sigma}$ it holds that $\sigma_i(h) = \sigma'_i(h)$. (This just means that, assuming all other parties play $\boldsymbol{\sigma}_{-i}$, play proceeds identically whether P_i plays σ_i or σ'_i.)

Definition 10. *Let Γ be an extensive form game, and let $\boldsymbol{\sigma}$ be a Nash equilibrium in Γ. Then $\boldsymbol{\sigma}$ is* stable with respect to trembles (for realizable histories) *if there exists an $\epsilon > 0$ such that for all i and every $\boldsymbol{\sigma}'_{-i}$ with $d(\boldsymbol{\sigma}_{-i}, \boldsymbol{\sigma}'_{-i}) < \epsilon$ there exists a σ'_i that is a best response to $\boldsymbol{\sigma}'_{-i}$ and such that σ_i and σ'_i yield equivalent play with respect to $\boldsymbol{\sigma}$.*

2.6 Cryptographic Considerations

In a cryptographic setting, it is natural to modify the way games are treated and the way various equilibrium notions are defined. We give an example of how this might be done for the specific case of parties running a protocol in the standard cryptographic sense, though it can be easily extended for more general scenarios (for examples, parties running a protocol and then taking some action as in Section 3, or parties who receive some initial input as in Section 4).

As usual in the cryptographic setting, we introduce a security parameter k provided to all parties at the beginning of the game. The action of a player P_i now corresponds to running an interactive Turing machine (ITM) M_i. This ITM M_i takes as input some current state and incoming messages from the other parties, and outputs the next message of player P_i along with updated state. The message m_i is then sent to all other parties (we are assuming here that communication is over a broadcast channel). We require M_i to run in probabilistic polynomial-time, which we take to mean that the next message function is computed in time polynomial in k. This definition allows M_i to run for an unbounded number of rounds and, if desired, we can additionally require that the expected number of rounds for which M_i runs is also polynomial.

Utility functions take the security parameter k as input, and are functions mapping transcripts of a protocol execution to the reals that can be computed in time polynomial in k. We stress that, as in extensive form games, utilities depend only on the "observable outcome" of the game play.

For the purposes of this section, we define a *computational game Γ* to be one in which the actions of each player correspond to the set of probabilistic polynomial-time ITMs, and where the utilities of each player are polynomial-time computable. We remark that we no longer need to consider mixed strategies, since a mixed strategy that can be implemented in polynomial time corresponds to a pure strategy (since pure strategies correspond to randomized ITMs).

An important difference between the cryptographic setting and the setting we have considered until now is that *now parties are assumed to be indifferent to negligible changes in their utilities.* For example:

Definition 11. *Let $\Gamma = (\{A_i\}, \{u_i\})$ be a computational game. A strategy vector $\boldsymbol{M} = (M_1, \ldots, M_n)$ is a* computational Nash equilibrium *if for all i and any* *probabilistic polynomial-time ITM M_i' there is a negligible function ϵ such that*

$$u_i(k, M_i', \boldsymbol{M}_{-i}) - u_i(k, \boldsymbol{M}) \leq \epsilon(k).$$

I am unaware of any result characterizing conditions under which Nash equilibria exist in computational games.

Subgame perfection and related notions. Execution of a protocol can naturally be regarded as an extensive form game. Extending equilibrium notions for extensive form games to the computational setting is, however, less obvious. For example, a first approach to extending the notion of subgame perfection to the computational setting would be to say that the strategy vector $\boldsymbol{\sigma}$ of the game Γ is subgame perfect if for all possible histories h of Γ, the strategy vector $\boldsymbol{\sigma}^h$ is a computational Nash equilibrium of the reduced game Γ^h. However, this ignores the probability with which history h is reached! On the other hand, it is unclear how to assign a probability to a non-realizable history. We are not aware of any definition of computational subgame perfection that deals with these issues.

A recent definition suggested by Kol and Naor [29] explicitly rejects the idea of "weighting" the utility of strategies according to the probability with which a given history is reached. Instead, informally, they require that conditioned on reaching *any* history that occurs with positive probability, players' strategies should remain in equilibrium. In their definition of a computational game, they allow players to use ITMs which run in time polynomial in $k + r$, where r is the number of rounds that have been played thus far. (Thus, the next-message function in their case may be viewed as a function from the entire history/transcript thus far to a next message, rather than from some internal state and a set of incoming messages to a next message, as defined above.) For lack of any better name, we refer to ITMs of this sort as running in *liberal* polynomial time and refer to the notion of t-resilient* equilibria for strategy vectors that remain in equilibrium even with respect to this stronger class of machines. Finally, we let $u_i(\cdot \mid h)$ denote the expected utility of P_i conditioned on history h. We now give the definition of Kol and Naor:[2]

Definition 12. *Let Γ be a computational game. A strategy vector $\boldsymbol{M} = (M_1, \ldots, M_n)$ is* computationally t-immune[3] *if for every history h realizable with respect to $\boldsymbol{M}$ and every i, the strategy vector $\boldsymbol{M}^h$ is a t-resilient* Nash equilibrium in the reduced game Γ^h. I.e., for every $\mathcal{C} \subset [n]$ with $|\mathcal{C}| \leq t$ and every liberal polynomial-time ITM $M_{\mathcal{C}}'$ there is a negligible function ϵ such that*

$$u_i(k, M_{\mathcal{C}}', \boldsymbol{M}_{-\mathcal{C}} \mid h) - u_i(k, \boldsymbol{M} \mid h) \leq \epsilon(k).$$

2.7 Critiques of Game Theory

Without going into much detail here, I will simply say that it is not at all clear whether game theory provides the "best" way of modeling interactions, both in

[2] One change we introduce is to condition on observable histories rather than on players' random coins (which may be private).

[3] Note that immunity refers to an entirely different concept in [1].

general as well as specifically in a cryptographic setting. (All of the critiques I mention here are well-known, and not in any way novel.) For starters, it is unclear the extent to which the behavior of most people can be modeled as rational. (*Social economists* study exactly this issue.) Even if we are willing to believe that people act rationally, it is not always clear when a protocol designer can assume any knowledge of their utilities.

Irrespective of the above, many of the solution concepts are unsatisfying. The notion of a Nash equilibrium is perhaps the most intuitively appealing one, but in cases where multiple Nash equilibria exist it is unclear which one the parties will settle on or even if they can agree to settle on one at all. Other notions have been introduced in an effort to distinguish among various Nash equilibria, but it seems that for every such notion there exists a game in which applying the notion goes against one's intuition. (See, e.g., [19, pp. 462–463] for an example in the context of iterated deletion of weakly dominated strategies where it is to one party's advantage to publicly burn their money.)

3 Implementing Mediators using Cryptography

As we have seen in Section 2.3, if parties are willing to assume the existence of a trusted mediator then they can potentially achieve certain equilibria that may be "preferable" to any of the available Nash equilibria. If a trusted mediator is *not* available, the question becomes: *to what extent can the parties themselves run a protocol in place of the mediator?*

This question was first explored in the economics community [14, 6, 18, 9, 42, 43, 4] (see [2] for a summary of these results), where researchers suggested "cheap talk" protocols by which parties could communicate amongst themselves to implement a correlated equilibrium. (As the terminology suggests, communication among the players is "cheap" in the sense that it costs nothing; it is also "worth nothing" in the sense that players are not "bound" to any statements they make; e.g., there is no legal recourse if someone lies). In the cryptography community, the question was first addressed by Dodis, Halevi, and Rabin [15].

3.1 Defining the Problem

Let us begin by defining the basic problem. (Other variants and extensions will be explored below.) We are given some n-party game $\Gamma = (\{A_i\}, \{u_i\})$ in normal form, along with a correlated equilibrium $\mathcal{M}$. We then define the extensive form game Γ_{CT} in which all parties hold a common security parameter k and first communicate in a "cheap talk" phase. The parties then play Γ, making their moves simultaneously (as always). Following the game-theoretic convention, all parties must play some action in Γ. (I.e., we do not allow player P_i to "abort" in Γ unless this is an action in A_i.) On the other hand, following the cryptographic convention we *do* allow players to abort (and refuse to send any more messages) during the cheap talk phase.

We make no assumptions regarding the exact communication model during the cheap talk phase. For now, however, we assume that colluding parties can communicate "out of band" throughout the entire game. (This assumption is removed in Section 3.4.) Thus, for now we focus on *ex post* correlated equilibria which are resilient to coalitions even when such communication is allowed.

A player's strategy in Γ_{CT} determines both the protocol it runs in the cheap talk phase as well as the action it plays in Γ. We may now define the basic goal:

Definition 13. *Let Γ be a game, and let $\mathcal{M}$ be an ex post t-resilient correlated equilibrium in Γ. Let Γ_{CT} be the cheap talk extension of Γ, and let $\boldsymbol{\sigma}$ be an efficient strategy vector in Γ_{CT}. Then $\boldsymbol{\sigma}$ is a* t-resilient implementation of $\mathcal{M}$ *if (1) $\boldsymbol{\sigma}$ is a t-resilient computational equilibrium in Γ_{CT}, and (2) for all i, it holds that $u_i(k, \boldsymbol{\sigma}) = u_i(k, \mathcal{M})$.*

One might strengthen the definition to require that the *distribution* of payoffs in Γ_{CT} (both for each party as well as when considering joint distributions among multiple parties) is close to the distribution of payoffs in the original mediated game. A stronger requirement of a different flavor is given by Lepinski et al. [30], who require (informally) that any vector of expected payoffs achievable by $\mathcal{C}$ in Γ_{CT} (i.e., even ones that are sub-optimal for $\mathcal{C}$) can also be achieved by $\mathcal{C}$ in the original mediated game. We do not impose such requirements here.

3.2 A Simple Observation

It is instructive to begin with a relatively simple observation: if t, n, and the communication model are such that *completely fair* secure multi-party computation [20, Def. 7.5.4] is possible, then *any* correlated equilibrium $\mathcal{M}$ of any game Γ has a t-resilient implementation: During the cheap talk phase the parties run a completely fair protocol Π computing $\boldsymbol{a} \leftarrow \mathcal{M}$, where P_i receives a_i as output. Following the cheap talk phase, each party plays the action it received as output in Π. It is not hard to see that the strategy vector thus specified (i.e., "run Π and then play the result") is a t-resilient (computational) equilibrium with expected payoffs identical to those in the original mediated game.

Applying the above observation to the standard communication model, we see that if parties are connected by pairwise point-to-point channels then a t-resilient implementation of any correlated equilibrium exists when $t < n/3$. If a broadcast channel or a PKI is additionally assumed, then t-resilient implementations exist whenever $t < n/2$. The above all follow from standard results in secure multi-party computation [11, 8, 40, 7]. Lepinski et al. [30] show how to achieve completely fair secure computation for any $t < n$ — and hence show t-resilient implementations of any correlated equilibrium for $t < n$ — in a non-standard communication model where "secure envelopes" are assumed. (Completely fair secure multi-party computation using point-to-point channels and broadcast is, in general, impossible for $t \geq n/2$ [13].) Assuming secure envelopes may be reasonable in some settings, but such envelopes seem impossible to realize (without assuming trusted parties) in a distributed setting such as the Internet.

3.3 Implementing Mediators without Completely Fair MPC

The natural next question is: when can a correlated equilibrium be implemented even though completely fair secure computation is ruled out? The initial result in this direction is due to Dodis, Halevi, and Rabin [15], who examine the case $t = 1, n = 2$. Before explaining their solution, we first introduce some terminology. Let Γ be a game in normal form. Then the *minimax profile* against player P_i is an action $\boldsymbol{a}_{-i} \in A_{-i}$ (or, more generally, in the product distribution $\times_{j \neq i} \Delta(A_i)$) minimizing $\max_{a_i \in A_i}\{u_i(a_i, \boldsymbol{a}_{-i})\}$. In other words, a minimax profile $\boldsymbol{a}_{-i}$ "punishes" P_i by giving P_i its lowest possible utility, assuming P_i plays a best response to the strategy of the other parties.

The basic idea of Dodis, Halevi, and Rabin is as follows: Let $\mathcal{M}$ be a correlated equilibrium in some two-party game Γ. In Γ_{CT}, the two parties run a protocol Π for computing $(a_1, a_2) \leftarrow \mathcal{M}$, where party P_i receives a_i as output. This protocol Π is "secure-with-abort" (cf. [20, Def. 7.2.6]), which informally means that privacy and correctness hold but fairness does not; in particular, we assume it is possible for P_1 to receive its output even though P_2 does not. After running Π, each party plays the action it received as output in Π; if P_2 does not receive output from Π then it plays the minimax profile against P_1.

It is not hard to see that this is a 1-resilient implementation of $\mathcal{M}$. First, it is immediate that if both parties play the indicated strategy, then the payoffs of both parties in Γ_{CT} are exactly the payoffs they would receive by playing $\mathcal{M}$ in Γ. Let us now argue that this is a computational Nash equilibrium in Γ_{CT}. We first observe that P_2 has no incentive to deviate: no matter how P_2 plays when running Π, party P_1 receives correctly-distributed output and plays according to the correlated equilibrium. Given this, P_2's best action to play in Γ is given by its own output from Π. We remark that here we are relying on the assumption that P_2 can only run *polynomial-time* strategies, and that P_2 is indifferent to negligible differences in its expected utility, exactly as we have defined things in Definition 11.

As for P_1, the only way it can (effectively) deviate during the cheap talk phase is by running Π until it receives its own output a_1 and then possibly aborting the protocol so that P_2 does not receive any output. We claim that it is never to P_1's advantage to abort. (Note that the analysis in [15] seems to assume that P_1 either *never* aborts or *always* aborts, but of course P_1 can determine whether to abort based on its output.) If P_1 allows P_2 to receive its output, this induces some mixed strategy σ_2 that will be played by P_2. (I.e., σ_2 represents the marginal distribution on P_2's recommended action according to $\mathcal{M}$, conditioned on the fact that P_1's recommended action is a_1.) Since $\mathcal{M}$ is a correlated equilibrium, a_1 is a best response to σ_2. If P_1 aborts, then P_2 will play a minimax profile σ_2' against P_1. By definition of a minimax profile, P_1's best response to σ_2' cannot give P_1 better utility than its best response to σ_2. We conclude that P_1 always does worse by aborting the execution of Π. Given that both parties receive output in Π, it is obviously to P_1's advantage to play its recommended action. This completes the proof.

Extensions. The above ideas do not extend easily to a more general setting. For example, consider the case $t = 1, n = 3$ (with point-to-point communication). If these parties run a protocol Π in which a single deviating party can abort the computation *without being identified*, then the remaining parties do not know which player to "punish". In fact, essentially this situation is inherent in general [2, Theorem 4]. On the other hand, specific correlated equilibria may be implementable using the general approach discussed below.

Next look at the case $t = 2, n < 5$. Observe that even if one party, say P_1, is identified as cheating, the naive approach of having the remaining parties play a minimax profile against P_1 may not work. For one thing, although such a profile might result in a worse payoff for P_1, it may actually lead to a better payoff for a second player, say P_2, colluding with P_1. (And recall that 2-resilience only holds if deviations help *no one* in the coalition.) Moreover, if players play a minimax profile against P_1, it may be possible for P_2 (who, recall, is colluding with P_1) to deviate from the minimax profile and thus benefit P_1.

We are thus motivated to define a stronger notion of "punishment". Following [1, Def. 5], though differing in some respects, we define:

Definition 14. *Let Γ be an n-party game with correlated equilibrium $\mathcal{M}$. A strategy vector $\boldsymbol{\sigma}$ is a t-*punishment strategy *with respect to $\mathcal{M}$ if for all $\mathcal{C} \subset [n]$ with $|\mathcal{C}| \leq t$, all $\boldsymbol{\sigma}'_{\mathcal{C}}$, and all $i \in \mathcal{C}$ it holds that $u_i(\boldsymbol{\sigma}'_{\mathcal{C}}, \boldsymbol{\sigma}_{-\mathcal{C}}) \leq u_i(\mathcal{M})$.*

That is, any coalition would be better off following the recommendations of $\mathcal{M}$ rather than playing against $\boldsymbol{\sigma}_{-\mathcal{C}}$.

If a t-punishment strategy is available for a given correlated equilibrium $\mathcal{M}$, then this gives hope that a variant of the Dodis-Halevi-Rabin approach can be used to give a t-resilient implementation of $\mathcal{M}$. See [1, 2] for work along these lines. Also relevant is the work of [31, 27, 28], discussed in more detail in the following section. As we have mentioned earlier, a partial converse [2] of the positive result just mentioned shows that, in general, if a t-punishment strategy is *not* available for a given correlated equilibrium $\mathcal{M}$, then this equilibrium cannot be implemented. Further work is need to better characterize the exact conditions under which a given correlated equilibrium can or cannot be implemented.

3.4 Implementing *ex ante* Equilibria (and More)

This section provides a brief discussion of work aimed at a slightly different aspect of the problem. Assume now that colluding parties *cannot* communicate "out of band" once Γ_{CT} begins; i.e., during the cheap talk phase of Γ_{CT} all communication is done over a public channel, and after the cheap talk phase — when it is time for the parties to play Γ — there is no inter-party communication at all. (Colluding parties can try to communicate over the broadcast channel, but if they are obvious about it then this will be detected by the other parties and punished.) It is then meaningful to ask whether it is possible to implement an *ex ante* correlated equilibrium of Γ in the cheap talk extension Γ_{CT}.

This problem is not immediately solved even if completely fair secure computation is possible. The problem is that *covert channels* may exist in the protocol itself. If such covert communication is possible, an ex ante correlated equilibrium may no longer remain an equilibrium. Informally, say a protocol is *collusion free* if covert communication is impossible. (We remark, however, that it seems sufficient here to prevent covert communication only *after* the parties have learned their output, since communication between the colluding parties before they learn their recommended actions will not affect an ex ante equilibrium.) Lepinski et al. [31] show how to construct a collusion-free protocol assuming the existence of "secure envelopes"; their work is further developed in [27, 28]. Some impossibility results for collusion-free protocols are shown in [31], though it is not clear what are the implications of these results for the specific problem of implementing ex ante correlated equilibria.

Collusion freeness may also be interesting in other contexts; see [31, 27, 28] for further discussion. Recent work [27, 28] has looked at stronger notions of collusion freeness, with the aim of achieving game-theoretic guarantees such as *strategic equivalence* between a mediated game and the cheap talk implementation of it. In that work, it is assumed that parties cannot communicate "out of band" *even before the protocol begins*; furthermore, a protocol should not only prevent covert communication between parties but should also prevent parties from agreeing on a common bit. We do not give further discussion here.

3.5 Future Directions

The immediate open question is to further characterize when a given *ex post* correlated equilibrium of a game is implementable (in, say, the standard communication model, either with or without broadcast). One direction to explore is when using a *partially fair* protocol [34, 17, 13, 12, 21] might suffice. Also, recent results [22] show that complete fairness for $t \geq n/2$ is achievable *for certain functions* in the standard communication model, thus giving hope that for certain restricted classes of correlated equilibria a cheap talk implementation might be possible even when general fair computation is not. Yet another direction is to explore other communication models, e.g., when a simultaneous broadcast channel is available. Or, taking a cue from the work on collusion-free protocols, we may ask what can be achieved under the assumption that colluding parties cannot communicate once the protocol begins. (Cleve's impossibility proof [13] fails in both the aforementioned settings.) These questions are interesting both in the current context as well as in a purely cryptographic sense.

In another direction, we can strengthen Definition 13 to require cheap talk protocols to satisfy stronger game-theoretic notions such as subgame perfection. (See also the following section.) The Dodis-Halevi-Rabin approach, in particular, will usually not yield a subgame perfect equilibrium.

4 Rational Multi-Party Computation

We now briefly discuss the second research direction mentioned in the Introduction. Here, there is no underlying game; rather, the protocol itself *is* the game, in the sense that parties' utilities are now functions of the inputs and outputs of the parties running the protocol. The difference between this setting and the standard setting of secure computation is that, in contrast to the standard setting where some parties are assumed to follow the protocol and other may behave arbitrarily, in the current setting we only guarantee that all players are *rational*. (Thus, the models are incomparable.) The questions here are: *how can we construct "meaningful" protocols in this setting?* and (more tantalizingly) *does this setting enable us to circumvent impossibility results that hold under the standard definition of secure computation?*

Let us jump right in with a "straw man" definition that, as far as I know, is new. Assume a set of parties $P_1, \ldots, P_n$ where party P_i begins holding input x_i. We assume the vector of inputs $\boldsymbol{x} = (x_1, \ldots, x_n)$ is chosen according to some known distribution $\mathcal{D}$. The parties want to compute a possibly probabilistic function f, where $f(\boldsymbol{x})$ outputs a vector $\boldsymbol{y} = (y_1, \ldots, y_n)$ and P_i receives y_i. The parties run some protocol $\boldsymbol{\Pi} = (\Pi_1, \ldots, \Pi_n)$, and we assume this protocol is *correct* in the sense that it yields the correct output if run honestly. (However, we do not assume the parties use their given inputs; see below.) The utility function of P_i is now a polynomial-time function of its view during the execution of $\boldsymbol{\Pi}$, the initial inputs $\boldsymbol{x}$, and the outputs $\boldsymbol{y}_{-i}$ of all other parties. (Note that inputs may be viewed as *types* in the sense defined in Section 2.) For treating coalitions, it seems best to define, for each possible coalition $\mathcal{C}$, a utility function $u_{\mathcal{C}}$ that is a function of the coalition's view, the inputs $\boldsymbol{x}$, and the outputs $\boldsymbol{y}_{-\mathcal{C}}$ of the other parties. We let Γ_{real} denote the real-world game thus defined.

In an ideal world computation of f (see [20]), a party P_i receiving input x_i can replace its input with some other value $x'_i = \delta_i(x_i)$; we allow δ_i to be probabilistic and allow $x'_i = \perp$, which is treated as an abort. After parties hand their inputs to the ideal functionality, the functionality computes $\boldsymbol{y} = f(\boldsymbol{x}')$ and gives y_i to P_i. Each party then outputs an arbitrary (polynomial-time) function $\pi_i(\cdot)$ of its view; this is left implicit in what follows, and we thus let δ_i stand for the entire strategy of P_i in the ideal world game Γ_f. The utility functions u_i are as above, except that these are now applied to the output of P_i, the inputs $\boldsymbol{x}$, and the outputs $\boldsymbol{y}_{-i}$ of the other parties (and analogously for coalitions).

Shoham and Tennenholtz [41] define the class of *NCC functions* for which, roughly speaking, setting δ_i to the identity function is a Nash equilibrium for all $\mathcal{D}$. Focusing on NCC functions appears to be a mistake that unnecessarily limits the class of functions under study.

Let $\Pi_i \circ \delta_i$ denote the real-world strategy where P_i changes its input x_i to $x'_i = \delta_i(x_i)$, and then runs Π_i using input x'_i. Then:

Definition 15. *Let $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)$ be a Nash equilibrium of Γ_f with respect to utilities $\{u_i\}$ and input distribution $\mathcal{D}$. Then $\boldsymbol{\Pi}$ is a* Nash protocol *for f (with respect to $\boldsymbol{\delta}$, $\{u_i\}$, and $\mathcal{D}$) if (1) $\boldsymbol{\Pi} \circ \boldsymbol{\delta} = (\Pi_1 \circ \delta_1, \ldots, \Pi_n \circ \delta_n)$ is a computational Nash equilibrium in Γ_{real}, and (2) for all i, it holds that $u_i(k, \boldsymbol{\Pi} \circ \boldsymbol{\delta}) = u_i(k, \boldsymbol{\delta})$.*

A definition of t-resilience may be derived from the above. Note that privacy, etc. are *not* explicitly required; it is our belief that questions of rationality should be separated from questions of security against malicious behavior.

An easy observation is that any protocol for completely fair secure computation tolerating t malicious parties is a t-resilient protocol for any $\boldsymbol{\delta}$, $\{u_i\}$, and $\mathcal{D}$. We also conjecture that if a protocol $\boldsymbol{\Pi}$ is resilient for *all* $\boldsymbol{\delta}$, $\{u_i\}$, and $\mathcal{D}$, then it is completely fair. Thus, things only become interesting if (1) we are in a setting where completely fair secure computation is impossible; and/or (2) we look at equilibrium concepts *stronger* than a Nash equilibrium. We briefly discuss these issues now. More extensive discussion will appear in the full version of this paper.

Constructing Nash protocols without completely fair MPC. This relates to the question, raised earlier, as to when relying on rationality of the parties might enable *circumvention* of impossibility results. As one example, depending on the utilities assumed it is possible to achieve complete fairness (which, note, is attained in the ideal model used in Definition 15) even in the presence of coalitions consisting of half or more of the parties [1, 35, 24, 29]. Similarly, it is possible to implement Byzantine agreement over point-to-point channels even in the presence of coalitions controlling 1/3 or more of the parties [23].

Rational secret sharing and stronger notions of equilibrium. Halpern and Teague [26] were the first to suggest that Nash protocols do not suffice but, instead, stronger notions are needed. As a motivating example [26], consider t-out-of-n secret sharing (here, $t < n$) under the assumption that each party (1) prefers to learn the secret above all else; and (2) otherwise, prefers other parties not learn the secret. Consider the naive protocol in which each party simply broadcasts their share. (We assume authenticated shares, so each party can choose either to broadcast the correct value or nothing.) This is clearly a Nash protocol, since no matter what any particular party does at least t parties broadcast their share and everyone reconstructs the secret. Nevertheless, it appears that each P_i would prefer *not* to broadcast: if at least t other parties broadcast, then everyone (including P_i) gets the secret as before; however, if fewer than t parties broadcast then only P_i recovers the secret. That is, following the protocol is weakly dominated by *not* following the protocol, and we might expect that no one follows the protocol. (and hence the protocol is not very useful).

To address this, Halpern and Teague suggest to look for Nash protocols where players' strategies survive iterated deletion of weakly dominated strategies. Such protocols were constructed in [26, 1, 35, 24].

Kol and Naor [29] argue that the requirement of surviving iterated deletion does not suffice to rule out protocols that are, intuitively, irrational. The notion is also difficult to work with and does not seem to capture intuition very well; moreover, it leads to other undesirable consequences such as the fact that, if we do not assume simultaneous channels (and thus allow rushing), then protocols in which two parties are supposed to speak in the same round are inherently problematic. (Since each party will simply wait for the other to go first.) Kol and Naor thus suggest another notion that we have given as Definition 12. Their definition rules out protocols that, intuitively, seem rational to follow.

We suggest to explore using the notion of resistance to trembles. (cf. Definition 10). This requirement rules out the naive protocol mentioned above as well as the counterexample of Kol-Naor; on the other hand, the protocols of [26, 1, 35, 24] appear to satisfy it.

The work of [27, 28] offers other definitions of rational MPC.

4.1 Future Directions

The community has not yet settled on a definition for rational MPC, and finding the "right" definition seems important for further progress in this area. Looking at constructions, we note that almost all positive results for rational MPC thus far assume the utility functions inherited from [26] (an exception is [23]); a natural step is to characterize when rational MPC is possible for other classes of utilities. One can also look for closer connections between the questions considered in Sections 3 and 4.

More broadly, one might explore applications of the ideas described here to scenarios that are more complicated than function evaluation; trust inference in distributed systems serves as one compelling example. Another direction is to realize that secure computation does not happen in a vacuum, but instead may occur within an existing legal framework; given this, game theory might be profitably applied to analyze protocols satisfying the definitions of [5, 32].

Acknowledgments. I am very grateful to Ran Canetti and the TCC '08 program committee for inviting me to write this survey, and to Ran for helpful discussions and suggestions.

References

1. I. Abraham, D. Dolev, R. Gonen, and J. Halpern. Distributed computing meets game theory: Robust mechanisms for rational secret sharing and multiparty computation. In *Proc. 25th ACM Symposium on Principles of Distributed Computing (PODC)*, pages 53–62, 2006.
2. I. Abraham, D. Dolev, and J. Halpern. Lower bounds on implementing robust and resilient mediators. In *5th Theory of Cryptography Conference (TCC)*, 2008. See also http://arxiv.org/abs/0704.3646.
3. R. Aumann. Subjectivity and correlation in randomized strategies. *J. Mathematical Economics*, 1:67–96, 1974.
4. R. Aumann and S. Hart. Long cheap talk. *Econometrica*, 71(6):1619–1660, 2003.
5. Y. Aumann and Y. Lindell. Security against covert adversaries: Efficient protocols for realistic adversaries. In *4th Theory of Cryptography Conference (TCC)*, 2007.
6. I. Barany. Fair distribution protocols or how the players replace fortune. *Mathematics of Operations Research*, 17(2):327–340, 1992.
7. D. Beaver. Multiparty protocols tolerating half faulty processors. In *Advances in Cryptology — Crypto '89*, pages 560–572.
8. M. Ben-Or, S. Goldwasser, and A. Wigderson. Completeness theorems for non-cryptographic fault-tolerant distributed computation. In *20th ACM Symposium on Theory of Computing (STOC)*, pages 1–10, 1988.
9. E. Ben-Porath. Cheap talk in games with incomplete information. *J. Economic Theory*, 108(1):45–71, 2003.

10. L. Buttyán and J.-P. Hubaux. *Security and Cooperation in Wireless Networks.* Cambridge University Press, 2007.
11. D. Chaum, C. Crépeau, and I. Damgård. Multiparty unconditionally secure protocols. In *20th ACM Symposium on Theory of Computing*, pages 11–19, 1988.
12. R. Cleve. Controlled gradual disclosure schemes for random bits and their applications. In *Advances in Cryptology — Crypto '89.*
13. R. Cleve. Limits on the security of coin flips when half the processors are faulty. In *18th ACM Symposium on Theory of Computing (STOC)*, pages 364–369, 1986.
14. V. Crawford and J. Sobel. Strategic information transmission. *Econometrica*, 50(6):1431–1451, 1982.
15. Y. Dodis, S. Halevi, and T. Rabin. A cryptographic solution to a game theoretic problem. In *Advances in Cryptology — Crypto 2006.*
16. Y. Dodis and T. Rabin. Cryptography and game theory. In Nisan et al. [38].
17. S. Even, O. Goldreich, and A. Lempel. A randomized protocol for signing contracts. *Comm. ACM*, 28(6):637–647, 1985.
18. F. Forges. Universal mechanisms. *Econometrica*, 58(6):1341–1364, 1990.
19. D. Fudenberg and J. Tirole. *Game Theory.* MIT Press, 1991.
20. O. Goldreich. *Foundations of Cryptography, vol. 2: Basic Applications.* Cambridge University Press, 2004.
21. S. Goldwasser and L. Levin. Fair computation of general functions in presence of immoral majority. In *Advances in Cryptology — Crypto '90*, pages 77–93.
22. S.D. Gordon, C. Hazay, J. Katz, and Y. Lindell. Complete fairness in secure two-party computation. Manuscript, 2007.
23. S.D. Gordon and J. Katz. Byzantine agreement with a rational adversary. Rump session presentation, Crypto 2006.
24. S.D. Gordon and J. Katz. Rational secret sharing, revisited. In *Security and Cryptography for Networks (SCN)*, pages 229–241, 2006.
25. J. Halpern. Computer science and game theory: A brief survey. In *The New Palgrave Dictionary of Economics.* Palgrave Macmillan, 2nd edition. Anticipated publication date: 2008. Available from http://www.cs.cornell.edu/home/halpern/.
26. J. Halpern and V. Teague. Rational secret sharing and multiparty computation. In *36th Annual ACM Symp. on Theory of Computing (STOC)*, pages 623–632, 2004.
27. S. Izmalkov, M. Lepinski, and S. Micali. Rational secure computation and ideal mechanism design. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, 2005. See also technical report MIT-CSAIL-TR-2007-040, available at http://hdl.handle.net/1721.1/38208.
28. S. Izmalkov, M. Lepinski, and S. Micali. Verifiably secure devices. In *5th Theory of Cryptography Conference (TCC)*, 2008.
29. G. Kol and M. Naor. Cryptography and game theory: Designing protocols for exchanging information. In *5th Theory of Cryptography Conference (TCC)*, 2008.
30. M. Lepinski, S. Micali, C. Peikert, and A. Shelat. Completely fair SFE and coalition-safe cheap talk. In *23rd Annual ACM Symposium on Principles of Distributed Computing (PODC)*, pages 1–10, 2004.
31. M. Lepinski, S. Micali, and A. Shelat. Collusion-free protocols. In *ACM Symposium on Theory of Computing (STOC)*, 2005.
32. A.Y. Lindell. Legally-enforceable fairness in secure multiparty computation. In *CT-RSA 2008.* to appear.
33. N. Linial. Game-theoretic aspects of computer science. In R. Aumann and S. Hart, editors, *Handbook of Game Theory with Economic Applications*, volume 2, pages 1340–1395. North Holland, 1994.

34. M. Luby, S. Micali, and C. Rackoff. How to simultaneously exchange a secret bit by flipping a symmetrically-biased coin. In *24th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 11–21, 1983.
35. A. Lysyanskaya and N. Triandopoulos. Rationality and adversarial behavior in multi-party computation. In *Advanced in Cryptology — Crypto 2006*.
36. H. Moulin and J.-P. Vial. Strategically zero sum games. *Intl. J. Game Theory*, 7(3/4):201–221, 1978.
37. J. Nash. Non-cooperative games. *Annals of Mathematics*, 54(2):286–295, 1951.
38. N. Nisan, T. Roughgarden, É. Tardos, and V. Vazirani, editors. *Algorithmic Game Theory*. Cambridge University Press, 2007.
39. M. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 2004.
40. T. Rabin and M. Ben-Or. Verifiable secret sharing and multiparty protocols with honest majority. In *21st ACM Symp. on Theory of Computing*, pages 73–85, 1989.
41. Y. Shoham and M. Tennenholtz. Non-cooperative computation: Boolean functions with correctness and exclusivity. *Theoretical Comp. Sci.*, 343(1–2):97–113, 2005.
42. A. Urbano and J. Villa. Computational complexity and communication: Coordination in two-player games. *Econometrica*, 70(5):1893–1927, 2002.
43. A. Urbano and J. Villa. Computationally restricted unmediated talk under incomplete information. *Economic Theory*, 23(2):283–320, 2004.