Incentive Compatibility in Two-Stage Repeated Stochastic Games

Bharadwaj Satchidanandan© and Munther A. Dahleh ©, Fellow, IEEE

Abstract—In this article, we address mechanism design for two-stage repeated stochastic games—a setting using which many emerging problems in electricity markets can be readily modeled. We introduce the notion of dominant strategy nonbankrupting equilibrium (DNBE) which requires players to make very few assumptions about the behaviors of other players in order to employ their equilibrium strategy. Consequently, a mechanism that renders truth-telling a DNBE could be quite effective in molding real-world behavior along truthful lines. We present a mechanism for two-stage repeated stochastic games that renders truth-telling a DNBE.

Index Terms—Dominant strategy nonbankrupting equilibrium (DNBE), incentive compatibility, mechanism design, repeated stochastic games.

I. INTRODUCTION

A CENTRAL problem in power system operation is to balance the demand and supply of power at all times. This has to be done in the presence of disturbances such as generation loss, load change, transmission line failures, etc. Such balance is achieved by a combination of feedforward and feedback control. The latter is performed by systems, such as turbine-governor controllers, load frequency controllers, power system stabilizers, etc. Feedforward control is done economically via markets. The system operator obtains in the day-ahead market the load forecast for the following day and accordingly plans the power dispatch of each generator by solving an economic dispatch problem. The generators are contractually obliged to supply the requisite power in real time. Given that power generators have largely been fossil-fuelled and therefore controllable, and that loads have largely been predictable, this approach has been viable so far.

Going forward, the power system is expected to undergo a revolution. The coming decade could witness increased renewable energy penetration, electric vehicle (EV) penetration, EV energy storage integration, Demand Response (DR), etc. The increased stochasticity introduces two challenges for electricity market operations which are elaborated upon next.

The first challenge pertains to optimal day-ahead planning. In the day-ahead market, a renewable power producer may not know the exact amount of power that it can produce the following day. Rather, it may only know a probabilistic description thereof. Similarly, a DR provider may not exactly know its baseline and its cost for curtailing power consumption the following day, but only know a probabilistic description thereof. Finally, an EV operator may not know the precise time intervals during which it can lease its battery to the grid the following day, but only know a probabilistic description thereof. It is only in real time that each of the aforementioned entities would come to know these quantities precisely. Consequently, it is not viable in the day-ahead market for a renewable power producer to commit to delivering a specified quantum of power in real time, for a DR provider to commit to curtailing a specified quantum of power in real time if called upon for DR, or for an EV operator to commit to leasing its battery for a specified duration the following day.

The best that they can do is provide probabilistic guarantees. This poses a challenge for day-ahead planning. Addressing it requires redesigning electricity markets to accommodate day-ahead commitments that are only probabilistic in nature.

A second and equally important challenge stems from the potential for strategic behavior by market participants. To elaborate, each market participant—whether it is a renewable power producer, a DR provider, or an EV operator—has associated with it a utility function which it seeks to maximize. The aforementioned stochasticities can be modeled by modeling the utility function of each market participant as being stochastic, whose probability distribution the participant knows in the day-ahead market, and whose realization the participant observes only in real time. Both the probability distribution and the realization of the utility function are in general private to the market participant. In particular, the system operator is not privy to this information, which it requires to compute the optimal day-ahead and real-time decisions. Consequently, the system operator requests each market participant to report in the day-ahead market the probability distribution of its utility function, and report in real time the realization of its utility function. However, being strategic, the participant may misreport either or both of these quantities precisely. Consequently, it is not viable in the day-ahead market for a renewable power producer to commit to delivering a specified quantum of power in real time, for a DR provider to commit to curtailing a specified quantum of power in real time if called upon for DR, or for an EV operator to commit to leasing its battery for a specified duration the following day.

Specifically, in a next-generation system, there is significant randomness in the power that a renewable generator can produce, the baseline of a DR provider, the cost that a DR provider incurs for curtailing power consumption, the time intervals during which an EV can lease its battery to the grid for energy storage, etc. The increased stochasticity introduces two challenges for electricity market operations which are elaborated upon next.
quantities if by doing so, there is a possibility for it to cause the system operator to take decisions that result in the accrual of a larger utility for itself, even if they deteriorate social welfare. See [1] and [2] for a more elaborate explanation. Hence, the system operator is confronted with the problem of eliciting both the probability distribution and the realization of each market participant’s utility function truthfully in order to operate the grid efficiently.

The two aforementioned challenges render the feedforward control mechanisms of today’s electricity markets ineffective in operating the next-generation power grid. New models and mechanisms must be devised to address a variety of such problems that arise in the context of next-generation power systems. To address this, we introduce the environment of two-stage repeated stochastic games using which many such problems can be readily modeled. The environment is an extension of the one-shot two-stage stochastic game introduced in [3] to repeated plays.

In a next-generation power system, the interactions between market participants (henceforth referred to as players) and the system operator on any particular day can be modeled as a two-stage stochastic game. Such a game, as the name suggests, consists of two stages. In the first stage, the players do not know their valuation functions precisely, but only know the probability distribution thereof. It is only in the second stage of the game that the valuation functions realize. However, the social planner cannot wait until the second stage to decide the outcome. It is constrained to make certain decisions in the first stage itself based on the probability distribution bids of the players’ valuation functions. Once the valuation functions realize and are reported in the second stage, the social planner can make corrections to the first stage decision by taking certain recourse actions, but this comes at a cost.

Motivated by applications to electricity markets which operate every day, we consider a setting wherein a two-stage stochastic game of the above form is played repeatedly, ad infinitum. Repeated playing affords the players a large class of strategies that adapt a player’s actions to all past observations and inferences obtained therefrom. In other settings, such as iterative auctions or dynamic games, where a large strategy space of this sort manifests, it typically has an important implication for mechanism design: it may be impossible to obtain truth-telling as a dominant strategy equilibrium (DSE) [4]. Consequently, in such scenarios, it is common to settle for mechanisms that render truth-telling only a Nash equilibrium (NE), or variants thereof, even though Nash equilibria are known to be poor models of real-world behavior. Among many reasons for this, a primary one is that each player has to make overly specific assumptions about the behaviors of other players in order to employ their NE strategy, which they may not do. In general, the lesser the burden of speculation in an equilibrium, the more plausible it is that it models real-world behavior.

Guided by the above maxim, we develop a new notion of equilibrium called dominant strategy nonbankrupting equilibrium (DNBE) that requires players to make very few assumptions about the behaviors of other players in order for them to employ their equilibrium strategy. Specifically, the only assumption that the players are required to make is that no player employs a strategy that leads to their own bankruptcy. We make this more precise in Section III. That this is a rather mild behavioral assumption requires no belaboring. Consequently, a mechanism that implements a certain desired behavior as a DNBE, as opposed to only a NE, could be quite effective in molding real-world behavior along the desired lines.

In this article, we present a mechanism for two-stage repeated stochastic games that renders truth-telling a DNBE. The mechanism is individually rational in that every player is guaranteed to accrue a nonnegative utility by truth-telling regardless of what strategies the other players employ. Finally, at equilibrium, social welfare is maximized. The mechanism is a generalization of the mechanism that we have developed in [1] for energy storage markets.

Finally, we illustrate an application of the proposed mechanism to DR markets. An aspect that we wish to highlight here is that there is a need to redesign the “bidding language” of electricity markets. In today’s electricity markets, the generators and the loads bid their supply and demand curves, respectively. However, with the inclusion of entities such as renewable generators, DR providers, and EVs, which all have significant stochasticities associated with their operation, the day-ahead market, unlike that of today, should allow for bids that are only probabilistic in nature. It is only in real time that these entities should be required to reveal their actual operating characteristics. The theory developed herein allows for such probabilistic bids to be submitted in electricity markets.

The rest of this article is organized as follows. Section II provides an account of related work. Section III presents a precise description of a two-stage repeated stochastic game, defines the notion of DNBE, and formulates the mechanism design problem. Section IV develops a mechanism for two-stage repeated stochastic games that guarantees truth-telling to be a DNBE. Section V describes the application of the results to the design of DR markets. Section VI concludes this article.

Notation: Vectors and sequences are denoted using boldface letters. Given a sequence \( x = \{x(1), x(2), \ldots\} \), we denote by \( x^l \) the segment \( \{x(1), \ldots, x(l)\} \). The hat notation is used to denote bids: given a variable \( x \) that is private to a player, we denote by \( \hat{x} \) the bid that the player submits for \( x \).

II. RELATED WORK

Ieong et al. [3] introduced the one-shot two-stage stochastic game environment and developed a mechanism that renders truthful bidding a sequential ex post NE. Tang and Jain [5] considered a two-stage game setting to model electricity markets consisting of wind power producers and develops a mechanism that incentivizes truthful bidding. However, it assumes that it is only in the first stage of the game that the wind power producers can bid strategically, and not in the second stage. In contrast, the setting that we have considered assumes that the valuation function distribution and the realization are private to the players, and that they can misreport either or both of them to accrue a higher utility. Such a setting could be relevant even in the context of designing markets for renewable power.
producers. Take for example a solar power producer. While on the one hand, it is true that the system operator can observe the realization of solar irradiance on any given day, the actual energy produced by a solar farm is not a function of only the irradiance. It could depend on a number of other local factors, such as the location of the farm, site-specific and time-specific shading, equipment quality, operations and maintenance, etc. These are factors that the system operator may not know or even wish to know. Consequently, a solar power producer could adjust its output power by adjusting, for example, the angle of the solar panels, and claim the resultant power to be the best that it can produce on that day. One could imagine similar issues arising in the context of distributed energy resources as well, such as rooftop solar on commercial buildings. While the veracity of a power producer’s claim can be checked by bestowing additional regulatory powers on the system operator, such a solution could be undesirable for a variety of reasons, especially when untruthful behavior can be kept in check by the market itself, as by the proposed mechanism. To account for such possibilities, the valuation function realizations of renewable power producers are also best modeled as being unobserved by the social planner.

Jackson and Sonnenschein [6] considered a Bayesian game wherein the distributions of private types of players are known to the social planner and a large number of independent copies of the game are concurrently played. The mechanism restricts the type bids of players across games to be consistent with their respective priors and shows that by a careful construction of the decision rule in each game, incentive compatibility can be ensured. While at a high level, the idea of this article is similar, there are two crucial differences. The first is that we do not assume the prior distributions of the players’ types to be known to the social planner. As mentioned before, this affords the players the prospect of jointly misreporting their type distributions and their realizations. Second and perhaps more important, the authors of [6] only address the case where the decisions for all games are taken simultaneously. On the other hand, if the games are played sequentially as in electricity markets, a richer strategy space emerges for the players as they can now adapt their actions at any time to all past observations and inferences obtained therefrom. We show that in spite of players having access to a richer class of history-dependent bidding strategies, which include strategies that potentially learn the type distributions and bidding strategies of other players, truth-telling can be rendered incentive compatible.

Mezzetti [7] presented a two-stage mechanism called the generalized Groves mechanism. In terms of the terminology and the framework presented in this article, the setting in [7] can be interpreted as each player having a privately known distribution of its valuation function which is required to be reported to the social planner. The joint distribution of the players’ valuation functions is assumed to be common knowledge. The social planner chooses an outcome that maximizes the expected social welfare based on the bids. After the social planner chooses the outcome, the valuation functions realize, which the players are required to bid in the second stage. Following this, a final payment is made. The payment rule guarantees that truth-telling by all players is an ex post NE. It is important to recognize that it is only the payment rule that has two stages in the aforementioned setting, and not the game itself. This in fact is one of the key departures of the one-shot two-stage stochastic game setting from the setting considered in [7]; the latter does not include the possibility for the social planner to take recourse actions after the valuation functions realize. In the context of electricity markets, not only is it feasible to take recourse actions but it also is imperative to take recourse actions if grid stability is to be maintained.

Kunimoto and Zhang [8] build upon [7] to devise a two-stage mechanism for bilateral trade. A power system offering a DR program is considered in [9] and [10] and a two-stage mechanism is presented using which a certain quantity of power can be apportioned among the loads when a DR event occurs. The first stage establishes a contingency plan that specifies the amount of power that would be supplied to each load in each contingency and the corresponding price, and the second stage, during which the contingency occurs, allows the loads to trade among themselves at the price established in the first stage. It is shown that the second-stage trade results in an allocation that pareto dominates the first-stage allocation.

All of the aforementioned papers consider a one-shot game, whereas the setting that we have considered is one of sequential playing. As mentioned before, the aspect of sequential playing introduces certain additional complexities for mechanism design that can be attributed to the availability of history-dependent bidding strategies to players—a challenge that pervades dynamic mechanism design theory. The authors in [11], [12], [13], [14], and [15] address the problem of mechanism design for dynamic games. The solution concept adopted in most of the literature on dynamic mechanism design is ex-post NE or variants thereof. With the exception of certain special cases, such as in [15], to the best of our knowledge, we are unaware of any work that tries to surpass NE or its variants and implement truth-telling in stronger notions of equilibria for broad classes of repeated or dynamic games, which is what we do in this article.

III. PROBLEM FORMULATION
A two-stage stochastic game played by \( n \) players and consisting of a social planner is described by

1. a publicly known set \( \Delta \) known as the type space of the players;
2. a publicly known set \( \Theta \) of probability mass functions over \( \Delta \), known as the supertype space of the players;
3. for each \( i \in \{1, \ldots, n\} \), a probability distribution \( \theta_i \in \Theta \), known as player \( i \)'s supertype, that is privately known to player \( i \) in the first stage of the game, and which it is supposed to report to the social planner in the first stage;
4. a set \( O_1 \) of first-stage outcomes;
5. a first-stage decision rule \( q_1^* : \Theta^n \rightarrow O_1 \) according to which the social planner chooses the first-stage outcome as a function of the players’ supertype bids;
6. for each \( i \in \{1, \ldots, n\} \), player \( i \)'s type \( \delta_i \in \Delta \) that is “drawn by nature” at random according to \( \theta_i \), whose
realization is privately observed by player \( i \) in the second stage of the game, and which it is supposed to report to the social planner in the second stage;
7) a set \( O_2 \) of second-stage outcomes or “recourse actions” that the social planner can choose;
8) a second-stage decision rule \( g_2^* : \Theta^n \times \Delta^n \to O_2 \) according to which the social planner chooses the second-stage outcome as a function of the players’ type and supertype bids;
9) a cost function \( c : \Theta_1 \times O_2 \to \mathbb{R} \) that specifies for every \((\sigma_1, o_2) \in \Theta_1 \times O_2\), the cost incurred by the social planner for choosing the outcome \( o_1 \) in the first stage and taking the recourse action \( o_2 \) in the second stage;
10) for each \( i \in \{1, \ldots, n\} \), a valuation function \( v_i : \Delta \times \Theta_1 \times \Theta_2 \to \mathbb{R} \) of player \( i \) that specifies for every \( \delta_i \in \Delta \) and every \((\sigma_1, o_2) \in \Theta_1 \times O_2\), the valuation of player \( i \) if its type is \( \delta_i \) and the social planner chooses the outcomes \( o_1 \) and \( o_2 \) in the first and the second stage of the game, respectively.

The first- and second-stage decision rules \((g_1^*, g_2^*)\) that we consider are those that maximize the expected social welfare. To elaborate, let \( g_1 : \Theta^n \to O_1 \) be any first-stage decision rule and \( g_2 : \Theta^n \times \Delta^n \to O_2 \) be any second-stage decision rule. If the players bid their types and supertypes truthfully, then the expected social welfare that results from the decision rule \((g_1, g_2)\) is

\[
E_{\delta - \theta} \left[ \sum_{i=1}^{n} v_i(\delta_i, g_1(\theta), g_2(\theta, \delta)) - c(g_1(\theta), g_2(\theta, \delta)) \right] =: W(\theta, g_1, g_2).
\]

The goal of the social planner is to maximize the expected social welfare, and so the decision rule \((g_1^*, g_2^*)\) that it employs is

\[
(g_1^*, g_2^*) = \arg\max_{g_1, g_2} W(\cdot, g_1, g_2)
\]

(1)

and the maximization is defined in the pointwise sense. The social planner computes \( g_1^* \) and \( g_2^* \) and announces it to the players before the game commences.

The problem that we study is one where a two-stage stochastic game of the above form is played repeatedly on each day \( l \), \( l \in \mathbb{Z}_+ \). For ease of exposition, we assume that the supertypes of the players remain the same on all days and that it is only their types that differ across days, although this assumption can be relaxed in a straightforward manner.\(^1\) Consequently, for each player \( i, i \in \{1, \ldots, n\} \), we denote by \( \theta_l \) its privately known supertype which remains the same on all days and by \( \delta_l \) its privately known type on day \( l \). The sequence \( \{\delta(1), \delta(2), \ldots\} \) is assumed to be independent and identically distributed (IID) with \( \delta(1) \sim \theta_1 \times \ldots \times \theta_n \).

A. First-Stage Strategy

On each day \( l \), each player \( i \) is required to report its supertype to the social planner in the first stage so that the latter can compute the optimal first-stage outcome. Since the players’ supertypes are assumed to remain the same on all days, it suffices for the players to bid their supertypes just once, namely, in the first stage of the game on day 1. Owing to strategic reasons that will become clear shortly, the players may not bid their supertypes truthfully, and so we denote by \( \theta_l \) the supertype bid of player \( i \) and by \( \sigma_l : \Theta \to \Theta \) the first-stage strategy according to which player \( i \) constructs its supertype bid. Therefore, \( \theta_l = \sigma_l(\theta_i) \). Once all players submit their supertype bids, the social planner computes the first-stage outcome as \( g_1^*(\sigma(\theta)) \), where \( \sigma(\theta) := [\sigma_1(\theta_1), \ldots, \sigma_n(\theta_n)] \). The game then proceeds to the second stage.

B. Second-Stage Bidding Policy

In the second stage on each day \( l \), each player \( i \) observes the realization of \( \delta_l(l) \) which it is supposed to report to the social planner. However, owing to strategic reasons that will become clear shortly, the players may not bid their type realizations truthfully, and so we denote by \( \hat{\delta}_l(l) \) player \( i \)’s type bid on day \( l \). We allow for the player to construct its type bid on any day using all information available to it until that day, and in accordance with any randomized history-dependent policy of its choosing. Specifically, a second-stage bidding policy \( \mu \) of player \( i \) is a rule which specifies for each \( \sigma_1 \in \Theta_1 \) and each \( l \in \mathbb{Z}_+ \), a probability distribution \( P_{\mu}(\hat{\delta}_l(l) \mid \sigma_1, \delta_l, \delta_{l-1}, \sigma_2, \ldots, \sigma_l, \delta_1 \) according to which player \( i \) constructs its second-stage bid \( \hat{\delta}_l(l) \) on day \( l \) if the first-stage outcome is \( \sigma_1 \). We denote by \( \Pi \), the set of all second-stage bidding policies available to player \( i \).

Note that the second stage bidding policy is a rule that maps the history of a player to its second-stage bid. While the outcome of the rule is random owing to the types and histories being random, there is nothing random about the rule itself. Consequently, a player without any loss of generality can choose its second-stage bidding policy right on day 1 as a function of its supertype—the only information that it has on day 1. This leads to the notion of a second-stage strategy which is described next.

C. Second-Stage Strategy

A second-stage strategy of player \( i \) is a function \( \pi_i : \Theta \to \Pi \), which specifies the second-stage bidding policy that it employs as a function of its private supertype \( \theta_i \), so that \( \pi_i(\theta_i) \) is the second-stage bidding policy of player \( i \).

Once all players submit their type bids, the social planner computes the second-stage outcome for day \( l \) as \( o_2(l) = g_2^*(\sigma(\theta), \hat{\delta}(l)) \).

Note that once the players’ first-stage and second-stage bidding strategies are fixed, a functional relationship is established between the types and the type bids, and all random variables become well defined.

D. Strategies and Strategy Profiles

We refer to the composition of the first- and second-stage strategies simply as a strategy, i.e., \( S_i := (\sigma_i, \pi_i) \) is referred to as the strategy of player \( i \). We denote by \( \Lambda_i \) the set of strategies
available to player $i$. We refer to $S := (S_1, \ldots, S_n)$ as the strategy profile of the players and denote by $\Lambda$ the set of strategy profiles $\Lambda_1 \times \ldots \times \Lambda_n$.

### E. Truthful Strategies

The stochasticity of the players’ types necessitates the definition of truthful strategy to be weaker than requiring the players to bid their types truthfully on all days. Such a notion is defined as follows.

**Definition 1:** A strategy $S_i = (\sigma_i, \pi_i)$ of player $i$, $i \in \{1, \ldots, n\}$, is truthful if

1. $\sigma_i(\theta) = \theta$ for every $\theta \in \Theta$; and
2. for every $\theta \in \Theta$ and every $\alpha_1 \in \mathcal{C}_1$, there exists $L \subseteq \mathbb{Z}_+$ with $\lim_{L \to \infty} 1/L \sum_{l=1}^L 1_{\{l \notin L\}} = 0$ such that for all $l \notin L$,
   \[
   \mathbb{P}_{\pi_i(\theta)}(\hat{\delta}_l(l)|\hat{\delta}_l^{-1}(l), \delta^{-1}_2, \alpha_1) = 1_{\{\delta_l(l) = \delta_l(l)\}}.
   \]

A strategy profile $(S_1, \ldots, S_n)$ is a truthful strategy profile if $S_i$ is truthful for every $i \in \{1, \ldots, n\}$.

In other words, a strategy $S_i$ is truthful if the supertype bid is truthful “almost all days.” We denote by $\mathcal{T}_i \subseteq \Lambda_i$, the set of all truthful strategies available to player $i$.

### F. Payments and Utilities

The social planner can pay or collect a monetary sum from each player at the end of each day. The payment that the social planner charges on any day can only depend on the information that is available to it until that day, namely, the supertype bids and the type bids submitted until that day. We denote by $p_{i,l}: \Theta_1 \times \ldots \times \Theta_n \times \Delta_1 \times \ldots \times \Delta_n \to \mathbb{R}$ the payment rule so that $p_{i,l}(\hat{\theta}, \hat{\delta}^l)$ specifies the amount that player $i$ is charged on day $l$, with a negative value denoting a transfer to the player. The utility accrued by player $i$ is defined as

\[
\begin{align*}
    u_i(S_i, S_{-i}, \theta, \delta^\infty) := \liminf_{L \to \infty} \frac{1}{L} \sum_{i=1}^L v_i \left( \delta_i(l), g_1(\hat{\theta}), g_2(\hat{\theta}, \hat{\delta}^l(l)) \right) - p_{i,l}(\hat{\theta}, \hat{\delta}^l).
\end{align*}
\]

(2)

Note that a player’s utility is a random variable that depends on the realization of the type sequence $\delta^\infty$.

### G. Dominant Strategy Nonbankrupting Equilibrium

As mentioned in Section I, a mild behavioral assumption, one that is quite likely to hold in practice, is that no player behaves in a manner that might result in its own bankruptcy. This is captured by the notion of a nonbankrupting strategy defined as follows.

**Definition 2:** A strategy $S_i$ of player $i$, $i \in \{1, \ldots, n\}$, is nonbankrupting if for all $(S_{-i}, \theta)$

\[
u_i(S_i, S_{-i}, \theta, \delta^\infty) > -\infty
\]

for all $\delta^\infty$, except perhaps on a set of probability zero.

A strategy profile $S = (S_1, \ldots, S_n)$ is nonbankrupting if $S_i$ is nonbankrupting for every $i \in \{1, \ldots, n\}$.
a nontruthful strategy than by employing a truthful strategy. This brings us to the mechanism design problem. The goal is to design a payment rule \( p_{i,t} : (i,t) \in \{1, \ldots, n\} \times \mathbb{Z}_+ \) such that a truthful strategy profile is a DNBE. The next section develops such a mechanism.

IV. EFFICIENT AND INCENTIVE-COMPATIBLE MECHANISM FOR TWO-Stage Repeated Stochastic Games

For each \( i \in \{1, \ldots, n\} \), the payment of player \( i \) on any day \( l \) consists of two components \( p^F_{i,t} \) and \( p^S_{i,t} \) that can be computed by the social planner at the end of the first and the second stage of the game, respectively, on day \( l \). These payment functions are defined next.

A. First-Stage Payment

The first-stage payment \( p^F_{i,t} \) is a function of only the first-stage bids of the players. Since the first-stage bids remain the same on all days, so do the first-stage payments. The first-stage payment is simply the Vickrey–Clarke–Groves (VCG) payment defined as

\[
p^F_{i,t} \left( \hat{\theta} \right) := W^* (\hat{\theta}_{-i}) - \mathbb{E}_{\delta \sim p_{-i}} \left[ \sum_{j \neq i} v_j \left( \delta_j \left( g^1_i (\hat{\theta}) \right), g^2_i (\hat{\theta}, \delta) \right) \right]
\]

\[
- c \left( g^1_i (\hat{\theta}), g^2_i (\hat{\theta}, \delta) \right)
\]

where \( \hat{\theta}_{-i} \) denotes the supertype bids of all players other than player \( i \).

B. Second-Stage Payment

At a high level, the first functional of the second-stage payment is to bind the first-stage and the second-stage strategies of the players. To achieve this, the second-stage payment rule compares the empirical frequencies of the players’ type bids with their supertype bids and penalizes discrepancies between them. To elaborate, denote by \( \hat{\theta}_i(t) \) the probability that a random variable distributed according to \( \hat{\theta}_i \) takes the value \( t, t \in \Delta \). On each day \( l \) and for each player \( i \), the second-stage payment rule computes the discrepancy

\[
\hat{f}_{i,t}(l) := \left[ \frac{1}{l} \sum_{t'=1}^{l} I_{[\hat{\theta}_i(t')=t]} \right] - \hat{\theta}_i(t)
\]

for every \( t \in \Delta \), and imposes a penalty of \( J_p(l) \) on player \( i \) if \( \hat{f}_{i,t}(l) \) falls outside a window of size \( r(l) \) for some \( t \), i.e., if

\[
|\hat{f}_{i,t}(l)| \geq r(l)
\]

for some \( t \in \Delta \).

In a repeated game, the sequence of second-stage outcomes serves as a source of common randomness which the players can potentially use to correlate their second-stage bids if there is a possibility for them to accrue a larger utility by doing so than by fabricating their bids independently of the other players’ bids. The second functionality of the second-stage payment is to disincentivize such strategies. Toward this end, on each day \( l \) and for each player \( i \), the second-stage payment rule computes

\[
\hat{h}_{i,d}(l) := \left[ \frac{1}{l} \sum_{t'=1}^{l} I_{[\hat{\theta}_i(t')=d, \hat{\theta}_{-i}(t')=d_{-i}]} \right] - \hat{\theta}_i(d_{-i})
\]

for every \( d \in \Delta^n \), and imposes a penalty of \( J_p(l) \) on player \( i \) if \( \hat{h}_{i,d}(l) \) falls outside a window of size \( r(l) \) for some \( d \in \Delta^n \), i.e., if

\[
|\hat{h}_{i,d}(l)| \geq r(l)
\]

for some \( d \in \Delta^n \).

How should the window sequence \( \{r\} \) be chosen? On the one hand, the window size \( r(l) \) must tend to zero as \( l \) tends to infinity for otherwise, the set of sequences \( \{\hat{\theta}\} \) that satisfy (7) and (9) would be “large,” thereby violating incentive compatibility. On the other hand, if \( \{r\} \) decays too quickly, then even truthful type bids would violate (7) and (9) infinitely often, thereby incurring a large penalty and violating individual rationality (IR). Hence, the sequence \( \{r\} \) should be chosen in a manner that balances the two objectives. This is achieved by choosing \( \{r\} \) such that

\[
\lim_{l \to \infty} r(l) = 0
\]

and for some \( \gamma > 0 \)

\[
r(l) \geq \sqrt{\frac{\ln 2}{2 \gamma}} \cdot \frac{l^{1+\gamma}}{l}
\]

for all \( l \in \mathbb{Z}_+ \).

To obtain an intuition for the condition (11), note that the empirical frequency \( \frac{1}{l} \sum_{t'=1}^{l} I_{[\hat{\theta}_i(t')=x]} \) resulting from the true type sequence of player \( i \) is a random variable with mean \( \theta_i(l) \) and standard deviation that scales as \( 1/\sqrt{l} \). Therefore, if the window size decays at the same rate, then the probability of the empirical frequency falling outside the window would remain at a constant value. This suggests that the window size must scale slower than at least \( 1/\sqrt{l} \). By scaling the window size only slightly slower than \( \sqrt{l} \), namely, the rate specified by (11), truthful bids are guaranteed to almost surely satisfy (7) and (9) for all but finitely many values of \( l \). This is established in Lemma 1.

How should the penalty sequence \( \{J_p\} \) be chosen? As shown in Lemma 1, truthful players incur a penalty only finitely often almost surely, and so the long-term average penalty that they incur is almost surely zero regardless of how the sequence \( \{J_p\} \) is chosen. Therefore, the only objective in the design of \( \{J_p\} \) is for every nontruthful strategy to incur a sufficiently high penalty. This is accomplished by choosing \( \{J_p\} \) to be any nonnegative sequence such that

\[
\lim_{l \to \infty} \frac{J_p(l)}{l} = \infty.
\]

2It suffices for (11) to hold only for all sufficiently large \( l \).
We now have the necessary quantities to define the second-stage payment function. Define the event
\[
E_{i,S}(l) := \left\{ \max_{t \in \Delta} |\bar{f}_{i,t}(l)| \geq r(l) \cup \max_{d \in \Delta^n} |\bar{h}_{i,d}(l)| \geq r(l) \right\}
\]
which denotes the occurrence of at least one of (7) and (9). The second-stage payment of player \(i\) on day \(l\) is defined as
\[
p_{i,l}(\hat{\theta}, \hat{\delta}) := \left[ v_i(\delta_i(l), g_1^*(\hat{\theta}), g_2^*(\hat{\theta}, \delta(l))) \right] - \mathbb{E}_{\delta \sim p_{\alpha}} \left[ v_i(\delta_i(l), g_1^*(\hat{\theta}), g_2^*(\hat{\theta}, \delta(l))) \right] + J_p(l)\mathbb{I}(E_{i,s}(l)).
\]
(14)

Note that if all players employ a truthful strategy, then the long-term average second-stage payment almost surely equals zero for every player.

The total payment \(p_{i,l}(\hat{\theta}, \hat{\delta})\) that player \(i\) transfers to the social planner on day \(l\) is
\[
p_{i,l}(\hat{\theta}, \hat{\delta}) = p_{i,l}^E(\hat{\theta}) + p_{i,l}^F(\hat{\theta}, \hat{\delta}).
\]
(15)

The following theorem establishes the incentive and optimality guarantees of the mechanism.

**Theorem 1:** Consider the two-stage repeated stochastic game induced by the payment rule (15).

1) A truthful strategy profile is a DNBE.
2) If for every \(i \in \{1, \ldots, n\} \) and every \(\theta\)
\[
W^*(\theta) - W^*(\theta_{-i}) \geq 0
\]
then every player obtains a nonnegative utility by employing a truthful strategy regardless of the strategies that the other players employ.
3) If every player employs a truthful strategy, then the long-term average social welfare (4) that results is almost surely equal to its optimal value \(W^*(\theta)\).

**Proof:** Arbitrarily fix \(\theta, \hat{\delta} \in \Lambda_i\) that player \(i\) employs, and the strategy profile \(S_{-i} \in \mathcal{NB}_{-i}\) that all other players employ. We begin with a lemma.

**Lemma 1:** For \(T_i \in \mathcal{T}_i\), it holds almost surely that
\[
\limsup_{L \to \infty} \frac{1}{L} \sum_{l=1}^L J_p(l)\mathbb{I}(E_{i,(T_i, S_{-i})}(l)) = 0.
\]
(17)

That is, if player \(i\) employs a truthful strategy, then the penalty that it pays is almost surely zero.

**Proof:** It suffices to show that \(\{E_{i,(T_i, S_{-i})}(l)\}\) almost surely occurs only finitely often. Arbitrarily fix \(\hat{\delta} \in \Delta^n\). Define \(F \) so that \(\mathbb{I}(\delta_1(l) = d_1, \delta_{-1}(l) = \hat{\delta}_{-1})\) is a martingale difference sequence bounded by unity. It follows from the Azuma–Hoeffding inequality that:
\[
P \left( \left\{ \frac{1}{L} \sum_{l=1}^L \mathbb{I}(\delta_{i}(l) = d_i) \left[ \mathbb{I}(\delta_{-i}(l) = d_{-i}) - \theta_i(d_i) \right] \geq r(l) \right\} \right) \leq 2e^{-2l^2r^2(l)}.
\]
(18)

Combining the above inequality with (11) implies
\[
P \left( \left\{ \frac{1}{L} \sum_{l=1}^L \mathbb{I}(\delta_{i}(l) = d_i) \left[ \mathbb{I}(\delta_{-i}(l) = d_{-i}) - \theta_i(d_i) \right] \geq r(l) \right\} \right) \leq \frac{1}{L^{1+\gamma}}.
\]
(19)

Using (8) and the fact that player \(i\) employs a truthful strategy, the above inequality implies
\[
P \left( |\hat{h}_{i,d}(l)| \geq r(l) \right) \leq \frac{1}{L^{1+\gamma}}.
\]
(20)

which in turn implies that \(\sum_{l=1}^\infty P(|\hat{h}_{i,d}(l)| \geq r(l)) < \infty\.

Invoking the Borel–Cantelli lemma, we have that \(|\hat{h}_{i,d}(l)| \geq r(l)\) almost surely occurs only finitely often.

Similarly, \(\mathbb{I}(\delta_{-i}(l) = d_{-i}, F_{l+1})\) is a martingale difference sequence bounded by unity and following the same sequence of arguments as above, it can be established that \(|\hat{f}_{i,d}(l)| \geq r(l)\) almost surely occurs only finitely often.

Since \(d\) is arbitrarily chosen, we have that for every \(d \in \Delta^n\), \(|\hat{h}_{i,d}(l)| \geq r(l)\) and \(|\hat{f}_{i,d}(l)| \geq r(l)\) almost surely occur only finitely often, and the desired result follows.

We have
\[
u_i(S_i, S_{-i}, \theta, \delta^\infty)
= \liminf_{L \to \infty} \frac{1}{L} \sum_{l=1}^L \left[ v_i(\delta_i(l), g_1^*(\hat{\theta}), g_2^*(\hat{\theta}, \delta(l))) \right] - p_{i,l}(\hat{\theta}, \hat{\delta})
\]
where \(\hat{\theta}\) and \(\hat{\delta}\) are determined in accordance with \(S\).

Substituting (5) and (14) into (15), substituting the resulting expression for \(p_{i,l}(l)\) into the above equality, and simplifying the result yields
\[
u_i(S_i, S_{-i}, \theta, \delta^\infty) = \left[ W^*(\hat{\theta}) - W^*(\hat{\theta}_{-i}) \right] + \left[ \liminf_{L \to \infty} \frac{1}{L} \sum_{l=1}^L \left( v_i(\delta_i(l), g_1^*(\hat{\theta}), g_2^*(\hat{\theta}, \delta(l))) \right) \right]
- \limsup_{L \to \infty} \frac{1}{L} \sum_{l=1}^L J_p(l)\mathbb{I}(E_{i,s}(l)).
\]
(17)

 Arbitrarily fix \(T_i \in \mathcal{T}_i\). Then, we obtain using Lemma 1 and some straightforward algebra that
\[
u_i(T_i, S_{-i}, \theta, \delta^\infty) - \nu_i(S_i, S_{-i}, \theta, \delta^\infty) = \left[ W^*(\hat{\theta}, \hat{\delta}_{-i}) - W^*(\hat{\theta}_{-i}) \right] + \limsup_{L \to \infty} \frac{1}{L} \sum_{l=1}^L \left( v_i(\delta_i(l), g_1^*(\hat{\theta}), g_2^*(\hat{\theta}, \delta(l))) \right)
- \liminf_{L \to \infty} \frac{1}{L} \sum_{l=1}^L J_p(l)\mathbb{I}(E_{i,s}(l)).
\]
\[
\begin{aligned}
+ \limsup_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} J_p(l) 1_{\{E_j, \psi(l)\}}.
\end{aligned}
\] (22)

In what follows, we show that the above quantity is almost surely nonnegative, implying that truthful strategy profiles are DNBE.

Define
\[
\nu_i(\tilde{\theta}_i, \tilde{\theta}_{-i}) := E_{(\tilde{\theta}_{-i}) \sim \tilde{\theta}_{-i}} \left[ v_i \left( \delta_i, g_1(\tilde{\theta}), g_2^*(\tilde{\theta}, \tilde{\delta}) \right) \right]
\]
(23)
and
\[
\nu_{-i}(\tilde{\theta}_i, \tilde{\theta}_{-i}) := E_{(\tilde{\theta}_{-i}) \sim \tilde{\theta}_{-i}} \left[ \sum_{j \neq i} v_j \left( \delta_j, g_1(\tilde{\theta}), g_2^*(\tilde{\theta}, \tilde{\delta}) \right) - c \left( g_1^*(\tilde{\theta}), g_2^*(\tilde{\theta}, \tilde{\delta}) \right) \right]
\]
(24)
so that
\[
W^* \left( \tilde{\theta}_i, \tilde{\theta}_{-i} \right) = \nu_i \left( \tilde{\theta}_i, \tilde{\theta}_{-i} \right) + \nu_{-i} \left( \tilde{\theta}_i, \tilde{\theta}_{-i} \right).
\] (25)

Let $\mu(\Delta_n)$ be the set of joint probability mass functions over $\Delta \times \Delta^n$. For $\psi \in \mu(\Delta_n)$, define
\[
\rho_i(\psi) := E_{(\tilde{\theta}_{-i}) \sim \tilde{\theta}_{-i}} \left[ v_i \left( \delta_i, g_1(\tilde{\theta}), g_2^*(\tilde{\theta}, \tilde{\delta}) \right) \right].
\]
(26)

Let $\Psi(\tilde{\theta}_i, \tilde{\theta}) \subset \mu(\Delta_n)$ be the set of joint probability mass functions with “$x$-marginal” distributed according to $\tilde{\theta}_i$ and “$y$-marginal” distributed according to $\tilde{\theta}_1 \times \ldots \times \tilde{\theta}_n$. Then, for every $\psi \in \Psi(\tilde{\theta}_i, \tilde{\theta})$, we have
\[
W^* \left( \tilde{\theta}_i, \tilde{\theta}_{-i} \right) \geq \rho_i(\psi) + \nu_{-i} \left( \tilde{\theta}_i, \tilde{\theta}_{-i} \right).
\] (27)

To see this, note that if $\langle \delta_i, \tilde{\theta}_{-i} \rangle \sim \tilde{\theta}_i \times \tilde{\theta}_{-i}$, then the social planner can map $\delta_i$ to a random variable $\tilde{\delta}_i$ using an appropriate probability transition kernel $P_{\tilde{\delta}_i | \delta_i}$ such that $(\tilde{\delta}_i, \tilde{\theta}_{-i}) \sim \psi \in \Psi(\tilde{\theta}_i, \tilde{\theta})$. Consequently, by choosing the first-stage outcome as $g_1^*(\tilde{\theta})$ and the second-stage outcome as $g_2^*(\tilde{\theta}, \tilde{\delta}_i)$, an expected social welfare of $\rho_i(\psi) + \nu_{-i} \left( \tilde{\theta}_i, \tilde{\theta}_{-i} \right)$ can be attained. It follows that the optimal expected social welfare $W^* \left( \tilde{\theta}_i, \tilde{\theta}_{-i} \right)$ is at least as large, which yields (27).

Suppose for a moment that each player $j \in \{1, \ldots, n\}$ employs a stationary second-stage bidding policy $\mu_j^0$ so that $\tilde{\delta}_j(l) \sim \tilde{\theta}_j$ is chosen as a function of $\delta_j(l)$ according to some probability kernel $P_{\tilde{\delta}_j | \delta_j}$ for every $l$. For player $j$’s strategy to be nonbankrupting, it is necessary that $\lim_{t \to \Delta} 1/L \sum_{l=1}^{L} 1_{\{ \tilde{\delta}_j(t) = \tilde{\theta}_j \}}$ almost surely for every $t \in \Delta$ for (7) would be violated infinitely often otherwise resulting in infinite average penalty. So, for every $j \in \{1, \ldots, n\}$, if player $j$’s strategy is to be nonbankrupting, then $P_{\tilde{\delta}_j | \delta_j}$ must be such that $\tilde{\delta}_j(1) \sim \tilde{\theta}_j$ given $\delta_j(1) \sim \theta_j$. It follows that for every $j \in \{1, \ldots, n\}$, $(\delta_j(1), \tilde{\delta}(1)) \sim \psi_j$ for some $\psi_j \in \Psi(\theta_j, \tilde{\theta})$. It also follows that $\{(\delta_j(1), \tilde{\delta}(1)), (\delta(2), \tilde{\delta}(2)), \ldots\}$ is a sequence of IID random variables, and so we obtain using the strong law of large numbers (SLLN) that the right-hand side (RHS) of (22) almost surely equals $W^* \left( \tilde{\theta}_i, \tilde{\theta}_{-i} \right) - W^* \left( \tilde{\theta}_i, \tilde{\theta}_{-i} \right) + [v_i(\tilde{\theta}_i, \tilde{\theta}_{-i}) - \rho_i(\psi_i)]$ Upon substituting (25), this becomes $W^* \left( \tilde{\theta}_i, \tilde{\theta}_{-i} \right) - \nu_{-i} \left( \tilde{\theta}_i, \tilde{\theta}_{-i} \right) - \rho_i(\psi_i)$, and combining it with (27) implies the nonnegativity of (22).

However, in order to fabricate the type bids, the players may not restrict just to stationary policies but can employ any history-dependent policy. The rest of the proof is devoted to showing that the same result, namely, the nonnegativity of (22), holds even in the general case where the players may employ any nonbankrupting strategy. The key to establishing this is the following lemma that characterizes the empirical joint distributions of the reported types when all players employ a nonbankrupting strategy.

Lemma 2: Suppose that for every $j \in \{1, \ldots, n\}$
\[
\limsup_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} J_p(l) 1_{\{E_j, \psi(l)\}} < \infty.
\] (28)
Then, for every $d \in \Delta^n$
\[
\lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} 1_{\{\delta(l) = d\}} = \Pi_{j=1}^{n} \tilde{\theta}_j(d_j).
\] (29)

Proof: It suffices to show that for all $d \in \Delta^n$ and all $k \in \{1, \ldots, n-1\}$
\[
\lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} 1_{\{\delta_k(l) = d_k \}} = \tilde{\theta}_k(d_k)
\]
(30)
and that
\[
\lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} 1_{\{\delta_k(l) = d_k \}} = \tilde{\theta}_k(d_k)
\]
(31)
where $d_k := [d_k d_{k+1} \ldots d_n]$ and $\tilde{\theta}_k(l)$ is defined likewise.

Combining (28) with (12) implies that $\limsup_{L \to \infty} \sum_{j=1}^{L} 1_{\{E_j, \psi(l)\}} < \infty$ for every $j \in \{1, \ldots, n\}$, i.e., the event sequence $\{E_j, \psi(l)\}$ occurs only finitely often. Hence, we obtain using (13) and (10) that for all $d \in \Delta^n$ and all $j \in \{1, \ldots, n\}$
\[
\lim_{L \to \infty} \hat{f}_j(d_j(L)) = 0
\] (32)
and
\[
\lim_{L \to \infty} \hat{f}_j(d_j(L)) = 0.
\] (33)
Substituting (6) in (32) implies
\[
\lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} 1_{\{\delta_j(l) = d_j \}} = \tilde{\theta}_j(d_j)
\] (34)
for all \( d_j \in \Delta \) and all \( j \in \{1, \ldots, n\} \), which in particular establishes (31).

Substituting (8) in (33) implies

\[
\lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} 1_{\{\delta_j(l) = d_j, \delta_j(l) = d_j\}} = \hat{\theta}_j(d_j) \left[ \lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} 1_{\{\delta_j(l) = d_j\}} \right]
\]

(35)

for all \( d \in \Delta^n \) and all \( j \in \{1, \ldots, n\} \). In concluding (35), we have assumed that the limit in the RHS exists, to justify which certain additional arguments are required. We omit these details since they might lessen the focus on the main aspects of the proof.

The equality (30) can now be established by noting that

\[
\lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} 1_{\{\delta_{k,n}(l) = d_{k,n}\}} = \lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} \sum_{t_{1:k-1}} 1_{\{\delta_1(l) = t_1, \ldots, \delta_{k-1}(l) = t_{k-1}, \delta_{k,n}(l) = d_{k,n}\}} = \sum_{t_{1:k-1}} \hat{\theta}_k(d_k) \times \left[ \lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} 1_{\{\delta_{k+1:n}(l) = d_{k+1:n}\}} \right]
\]

(36)

where the third equality follows from (35).

It follows from (12) that \( \lim \sup_{L \to \infty} 1/L \sum_{l=1}^{L} J_p(l) 1_{\{E_1 \leq \delta(l)\}} \) can only take values 0 and \( \infty \). In the latter case, the nonnegativity of (22) is immediate. In the former case, since \( S_{-i} \) is a nonbankrupting strategy profile, we have that for all \( j \in \{1, \ldots, n\} \),

\[
\lim \sup_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} J_p(l) 1_{\{E_1 \leq \delta(l)\}} < \infty
\]

(37)

almost surely. Consequently, Lemma 2 applies, and we get

\[
\lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} \psi_i(\delta(l)) = \hat{\theta}_i \left( \lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} J_p(l) 1_{\{E_1 \leq \delta(l)\}} \right)
\]

(38)

Now, consider the empirical joint distribution \( \psi_L(d, \hat{\delta}) := \frac{1}{L} \sum_{l=1}^{L} 1_{\{\delta(l) = d, \hat{\delta}(l) = \hat{d}\}} \), where \( d \in \Delta \) and \( \hat{d} \in \Delta^n \). Note that \( \psi_L \in \mu(\Delta, \Delta^n) \) for all \( L \in \mathbb{Z}_+ \). It follows from SLLN that for any \( d \in \Delta \), \( \lim_{L \to \infty} \sum_{l=1}^{L} \psi_L(d, \hat{\delta}) = \theta(d) \). Since (37) holds, we obtain using Lemma 2 that for any \( \hat{d} \in \Delta^n \),

\[
\lim_{L \to \infty} \sum_{l=1}^{L} \psi_L(d, \hat{\delta}) = \Pi_{1:n} \hat{\theta}_j(d_j), \ i.e., \ the \ sequence \ \{\psi_L\} \ \text{of empirical joint distributions is such that its x-marginal}
\]

approaches the distribution \( \theta \) and its y-marginal approaches the distribution \( \hat{\theta}_1 \times \ldots \times \hat{\theta}_n \). It can be shown as a consequence that \( \{\psi_L\} \) approaches the set \( \Psi(\theta, \hat{\theta}) \) in that

\[
\lim_{L \to \infty} \inf_{\psi \in \Psi(\theta, \hat{\theta})} \| \psi - \psi_L \| = 0 \quad \text{as} \quad L \to \infty, \quad \text{where} \quad \| \cdot \| \text{can be any norm defined on the set} \ \mu(\Delta, \Delta^n) \ \text{of also, the function} \ \rho_i : \mu(\Delta, \Delta^n) \to \mathbb{R} \ \text{defined in (26) is a continuous function over a compact set, and hence uniformly continuous. It follows that:}
\]

\[
\lim_{L \to \infty} \inf_{\psi \in \Psi(\theta, \hat{\theta})} \rho_i(\psi) \leq \sup_{\psi \in \Psi(\theta, \hat{\theta})} \rho_i(\psi).
\]

(39)

Note also that

\[
\frac{1}{L} \sum_{l=1}^{L} v_i(\delta(l), g_1^*(\theta), g_2^*(\hat{\theta}, \hat{\delta}(l))) = E_{v_i(\delta, g_1^*(\theta), g_2^*(\hat{\theta}, \hat{\delta}))} = \rho_i(\psi_L). \quad \text{Taking}
\]

the limit as \( L \to \infty \) and using (39) implies

\[
\lim_{L \to \infty} \inf_{\psi \in \Psi(\theta, \hat{\theta})} \rho_i(\psi) \leq \sup_{\psi \in \Psi(\theta, \hat{\theta})} \rho_i(\psi).
\]

(40)

Substituting (38) and (40) in (22) yields

\[
u_i(T_1, S_{-i}, \theta, \hat{\theta}_{-i}) \geq W^*(\theta, \hat{\theta}_{-i}) - W^*(\hat{\theta}_i, \hat{\theta}_{-i}) - \rho_i(\psi).
\]

Upon substituting (25), the RHS of the above inequality becomes \( W^*(\theta, \hat{\theta}_{-i}) - W^*(\hat{\theta}_i, \hat{\theta}_{-i}) - \rho_i(\psi) \). Combining this with (27) implies its nonnegativity, thereby establishing the nonnegativity of (22).

We now prove the second statement of the theorem. Arbitrarily fix \( S_{-i} \in \Delta_{-i} \) and \( T_i \in \mathcal{T}_i \). Using (15), (2) and Lemma 1, we obtain almost surely that \( \nu_i(T_1, S_{-i}, \theta, \hat{\theta}_{-i}) = [W^*(\theta, \hat{\theta}_{-i}) - W^*(\hat{\theta}_i, \hat{\theta}_{-i})] \geq 0 \), where the inequality follows from (16). Hence, truth-telling is individually rational for every player.

That the mechanism maximizes social welfare under truthful bidding is a straightforward consequence of the optimality of the first- and the second-stage decision rules.

Having established the theoretical guarantees of the mechanism, let us briefly turn our attention to some practical considerations.

One of the aspects of the mechanism that might appear to limit its real-world applicability is the requirement for players to know the entire probability distribution of their types, which could be too much information. However, in practice, it would likely be computers that do the bidding on behalf of market participants rather than actual human beings. It is well within the reach of such systems to learn over time complicated quantities such as probability distributions to any precision. Consequently, it may not be as demanding a requirement as it initially appears.
for bidders to know what their probability distributions are, and for them to report (or strategically misreport) them to the social planner. Having said that, it may be of interest to adapt the mechanism to require only approximate probability distributions, described for example in terms of a few moments, rather than exact distributions.

Second, there may be certain scenarios in which players have outside options that guarantee them a certain utility. In such cases, a more general IR constraint must be guaranteed that assures each player a utility that is at least as large as the outside option. Designing mechanisms that satisfy the more general IR constraint while simultaneously being budget balanced (or at least the least budget imbalanced if balance is not possible) is an important avenue for future research.

Finally, certain aspects of the mechanism, such as the constraints on the penalty sequence and the window size sequence, may appear too stringent for real-world applications. However, as true of most theories, the mechanism is not meant to be implemented as such. It is judicious approximations of the mechanism, carefully tailored to the implementation contexts, that are likely to get deployed.

V. APPLICATION TO DEMAND RESPONSE MARKETS

As mentioned in Section I, one of the motivating reasons for introducing the environment of a two-stage repeated stochastic game is its ability to readily model many problems that arise in the context of next-generation electricity markets. We illustrate one such problem in this section, namely, mechanism design for DR markets. In addition to illustrating an application of the proposed framework, the results of this section also serve to illustrate the benefits of using the proposed mechanism as opposed to other “natural” mechanisms that a policy-maker might employ in such scenarios.

One of the main requirements of power systems operations is that the power supply has to equal the random demand at each time. In conventional systems, the supply can be controlled, and so the generation is continuously adjusted to follow the random demand. However, at deep levels of renewable energy penetration, the generation also becomes random and uncontrollable. A popular paradigm for maintaining demand-supply balance in such a scenario is to make the demand follow the random supply. This is referred to as DR and is achieved by using incentives to modulate the consumption.

One of the key challenges in implementing DR is that in order to optimally allocate a desired consumption reduction among DR providers, their costs for curtailing consumption must be known, which are in general random and private to the loads, and which they could misreport to achieve more favorable allocations for themselves. The goal of the mechanism designer is to elicit both the probability distribution and the realization of the private costs truthfully. See [2] for more details. In what follows, we describe how the mechanism developed in the previous section can be applied to this problem.

A second and equally important challenge in DR pertains to private and random baselines of DR providers which could be strategically manipulated. However, in this paper, we restrict attention to a special case wherein the baselines are assumed to be either known to the system operator or truthfully reported, and it is only the costs for reducing consumption that are private to the loads. See [2] for the extension to the more general case wherein the baselines are also private knowledge and subject to strategic manipulation.

In this section, we overload certain notation. Specifically, whenever a DR market-specific quantity maps to a two-stage repeated stochastic game-specific quantity, the former will be denoted using the same symbol as the latter.

Consider a system consisting of n DR providers and a reserve generator. Each DR provider has a cost function that specifies the cost it incurs as a function of its power consumption reduction. We assume that the cost function is parameterizable and denote by δi(l) the parameter that specifies the cost function of DR provider i on day l. Hence, cDR(x, δi(l)) denotes the cost that DR provider i incurs on day l for curtailing its consumption by x units from its baseline. The sequence δ∞ is IID with δ(1) ∼ θ := θ1 × · · · × θn, where θi denotes the probability distribution of δi(1). The reserve generator has associated with it a production function cs : R → R which specifies the cost it incurs as a function of the power that it produces.

Denote by d(l) the power shortage on day l. The system operator wishes to minimize the social cost of compensating for the shortage and therefore wishes to determine the consumption reduction x∗(δ(l)) of the DR providers and the reserve generation g∗(δ(l)) on day l as

\[\arg\min_{x_1,\ldots,x_n,\theta} \sum_{i=1}^{n} c_{DR}(x_i, \delta_i(l)) + c_s(g_s)\]

subject to \[\sum_{i=1}^{n} x_i + g_s = d(l).\]

The problem of course is that the system operator does not know \{δ1(l), . . . , δn(l)\}, and so it requests the DR providers to bid their cost functions. Denote by ˜δ(l) the parameter that DR provider i bids on day l. The system operator computes x∗(δ(l)) and pays each DR load i a payment p(l) on day l for reducing its consumption by x∗ i (δ(l)). The average utility that DR provider i accrues is defined as

\[u^x_i := \lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} p(l) - c(x^*_i(\tilde{\delta}(l)), \delta_i(l)).\]

It is straightforward to see that the average utility of each DR provider is a function of not only its own bidding strategy but also the bidding strategy of the other DR loads. Consequently, a DR provider may not bid its cost truthfully if there is a possibility for it to obtain a larger utility by misreporting its cost. This in turn could result in the DR program operating in a manner that is not social cost-minimizing. This motivates the mechanism design problem. The mechanism presented in the previous section can be used to design the payments of DR providers. Theorem I then assures the incentive compatibility of truthful bidding of costs for the DR providers.

For our numerical study, we have taken c(x, δi) = δi/2x2, c_s(x, δ_s) = δ_s/2x^2, θ to be a product of beta distributions of
unit mean and variance 2, and $\delta_s(l)$ to also be beta distributed with the same parameters.

Fig. 2 illustrates how the payment resulting from the proposed mechanism behaves from the point of view of a randomly chosen DR provider. Specifically, we fix the cost function of a randomly chosen DR provider and plot how its average payment varies with the mean of the costs of the other DR providers. Qualitatively, the higher the mean cost of a DR provider, the higher the inelasticity of its demand. Hence, Fig. 2 quantifies the rate at which the payment received by a given DR load increases as the demands of the other loads become more inelastic.

An arguably natural alternative for the proposed mechanism is the posted price mechanism, wherein the system operator announces the payment $p_{pp}$ that the DR providers would receive per unit reduction in their power consumption. Each DR provider $i$ then chooses its curtailment $x^*_i,pp(l)$ on day $l$ as $x^*_i,pp(l) = \arg\min_{x_i} c_{DR}(x, \delta_i(l)) - p_{pp}x$. The residual mismatch $d(l) - \sum_{i=1}^n x^*_i,pp(l) = g_s(l)$ is purchased in the spot market at price $c_s(x, \delta_s(l)) = \frac{\delta_s(l)}{2} g_s^2(l)$. Such a mechanism has been employed, for example, in a prior DR trial in the United Kingdom.

How do such “simple” and “natural” alternatives compare with the proposed mechanism? Fig. 3 compares the social cost attained by the proposed mechanism with the social cost attained by the posted price mechanism. Two important observations are in order. First, note that there exists a price at which the posted price mechanism attains its minimum social cost. However, this price is a function of the type distributions of the DR loads, which are their private knowledge. This necessitates the system operator to perform price discovery to compute the optimal price—a process which is vulnerable to strategic manipulation by DR providers. Second, even if DR providers do not manipulate price discovery, the minimum social cost that can be attained by the posted price mechanism is no lower than what can be attained by the proposed mechanism. This of course should come

![Fig. 2](image1.png)

**Fig. 2.** Average payment received by a fixed DR provider as a function of the mean of the supertypes of the other DR providers. The fixed DR provider has cost parameter $\delta_i(l) = 4$ for all $l$, and the supertypes of the other loads are beta distributed with varying mean and a fixed variance of 2. Hence, the average payment received by a given load increases as the demands of the other loads become more inelastic.

![Fig. 3](image2.png)

**Fig. 3.** Social cost attained by the posted price mechanism versus the price.

as no surprise given the optimality guarantee of the proposed mechanism.

VI. CONCLUSION

In this article, we have considered two-stage repeated stochastic games, wherein private information is revealed over two stages and the social planner is constrained to make a decision in each stage. The setting models many important problems that arise in the context of next-generation electricity markets. Recognizing the limitation of Nash equilibria in molding real-world behavior, we have introduced the notion of a DNBE which requires players to make very few assumptions about the behaviors of other players to employ their equilibrium strategy. Consequently, a mechanism that implements a certain desired behavior as a DNBE could effectively mold real-world behavior along the desired lines. We have developed a mechanism for two-stage repeated stochastic games that implements truth-telling as a DNBE. The mechanism is also individually rational and maximizes social welfare.

REFERENCES

[1] B. Satchidanandan and M. A. Dahleh, “An efficient and incentive-compatible mechanism for energy storage markets,” *IEEE Trans. Smart Grid*, vol. 13, no. 3, pp. 2245–2258, May 2022.

[2] B. Satchidanandan, M. Roozbehani, and M. A. Dahleh, “A two-stage mechanism for demand response markets,” *IEEE Control Syst. Lett.*, vol. 7, pp. 49–54, Jun. 2023.

[3] S. Ieong, A. M.-C. So, and M. Sundararajan, “Stochastic mechanism design,” in *Proc. Int. Workshop Web Internet Econ.*, 2007, pp. 269–280.

[4] D. Bergemann and J. Välimäki, “Stochastic mechanism design,” in *Proc. Int. Workshop Web Internet Econ.*, 2007, pp. 235–239, 2009.

[5] W. Tang and R. Jain, “Market mechanisms for buying random wind,” *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1615–1623, Oct. 2015.

[6] M. O. Jackson and H. F. Sonnenschein, “Overcoming incentive constraints by linking decisions,” *Econometrica*, vol. 75, no. 1, pp. 241–257, 2007.

[7] C. Mezzetti, “Mechanism design with interdependent valuations: Efficiency,” *Econometrica*, vol. 72, no. 5, pp. 1617–1626, 2004.

[8] T. Kunimoto and C. Zhang, “Efficient bilateral trade with interdependent values: The use of two-stage mechanisms,” *School Econ.*, Singapore Manage. Univ., Singapore Econ. Statist. Work. Paper no. 14-2020, May 2020.

[9] J. A. Doucet, K. J. Min, M. Roland, and T. Strauss, “A two-stage mechanism to improve electricity rationing,” *Can. J. Econ./Revue Canadienne d’Economique*, vol. 29, pp. 8270–8275, 1996.
[10] J. A. Doucet, K. Jo Min, M. Roland, and T. Strauss, “Electricity rationing through a two-stage mechanism,” *Energy Econ.*, vol. 18, no. 3, pp. 247–263, 1996. [Online]. Available: https://www.sciencedirect.com/science/article/pii/014098839600014X

[11] A. Pavan, I. Segal, and J. Toikka, “Dynamic mechanism design: A Myersonian approach,” *Econometrica*, vol. 82, no. 2, pp. 601–653, 2014.

[12] D. P. Baron and D. Besanko, “Regulation and information in a continuing relationship,” *Inf. Econ. Policy*, vol. 1, no. 3, pp. 267–302, 1984.

[13] M. Battaglini, “Long-term contracting with Markovian consumers,” *Amer. Econ. Rev.*, vol. 95, no. 3, pp. 637–658, 2005.

[14] D. Bergemann and J. Valimaki, “Efficient dynamic auctions,” Cowles Foundation for Res. Econ., Yale Univ., New Haven, CT, USA, Cowles Found. Discussion Paper no. 1584, 2006.

[15] K. Ma and P. R. Kumar, “Incentive compatibility in stochastic dynamic systems,” *IEEE Trans. Autom. Control*, vol. 66, no. 2, pp. 651–666, Feb. 2021.

Bharadwaj Satchidanandan received the Ph.D. degree in electrical and computer engineering from Texas A&M University, College Station, TX, USA, where he was advised by Prof. P. R. Kumar. He is currently a Postdoctoral Researcher with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA, where he is hosted by Prof. Munther A. Dahleh. His research interests include cyber-physical systems, security, renewable energy, game theory, mechanism design, millimeter-wave networks, control theory, and communications.

Dr. Satchidanandan was the recipient of the Best Student Paper Award at the 9th International Conference on Communication Systems and Networks, a finalist for the Best Student Paper Award at the 6th Indian Control Conference, and a Graduation Day Presenter at the 2018 Information Theory and Applications Workshop. He was a runner-up in the global Hackathon contest conducted as a part of the 2017 Global Conference on Cyber Space, for which he was the recipient of an award from the Prime Minister of India.

Munther A. Dahleh (Fellow, IEEE) received the B.S. degree in electrical engineering from Texas A&M University, College Station, TX, USA, in 1983, and the Ph.D. degree in electrical and computer engineering from Rice University, Houston, TX, USA, in 1987. He has been with the Department of Electrical Engineering and Computer Science (EECS), Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, where he is currently the William A. Coolidge Professor of EECS, since 1987. He is also a Faculty Affiliate with the Sloan School of Management. He is the founding Director of the MIT Institute for Data, Systems, and Society. He leads a research group that focuses on decisions under uncertainty. His recent work focused on understanding the economics of data as well as deriving a foundational theory for data markets. His research interests include networked systems, information design, and decision theory with applications to social and economic networks, financial networks, transportation networks, neural networks, the power grid, and in learning high-dimensional unstructured stochastic dynamic systems from finite data.

Dr. Dahleh is a Fellow of International Federation of Automatic Control.