Multi-unit Double Auctions: Equilibrium Analysis and Bidding Strategy using DDPG in Smart-grids

Sanjay Chandlekar
International Institute of Information Technology (IIIT), Hyderabad, India
sanjay.chandlekar@research.iiit.ac.in

Easwar Subramanian
TCS Innovation Labs
Hyderabad, India
easwar.subramanian@tcs.com

Sanjay Bhat
TCS Innovation Labs
Hyderabad, India
sanjay.bhat@tcs.com

Praveen Paruchuri
International Institute of Information Technology (IIIT), Hyderabad, India
praveen.p@iiit.ac.in

Sujit Gujar
International Institute of Information Technology (IIIT), Hyderabad, India
sujit.gujar@iiit.ac.in

ABSTRACT
We present a Nash equilibrium analysis for single-buyer single-seller multi-unit k-double auctions for scaling-based bidding strategies. We then design a Deep Deterministic Policy Gradient (DDPG) based learning strategy, DDPGBBS, for a participating agent to suggest bids that approximately achieve the above Nash equilibrium. We expand DDPGBBS to be helpful in more complex settings with multiple buyers/sellers trading multiple units in a Periodic Double Auction (PDA), such as the wholesale market in smart-grids. We demonstrate the efficacy of DDPGBBS with Power Trading Agent Competition’s (PowerTAC) wholesale market PDA as a testbed.

KEYWORDS
Multi-unit Periodic Double Auction; Equilibrium Analysis; Bidding Strategy; DDPG for Learning Equilibrium Strategy

1 INTRODUCTION
A double auction is a process of buying or selling goods or items [21] involving multiple buyers and sellers placing their bids/asks. It is extensively used to trade stocks, energy, and many other goods and services in the real world [7, 10]. The double auction plays a significant role in smart-grids [2], where multiple power generating companies (GenCos) and energy brokers trade energy in the wholesale market through PDA. PDA is a specific type of double auction where bids are cleared periodically in a sequence of pre-defined time periods. As PDA involves multiple discrete clearing periods, the buyer needs to participate in a series of auctions, and therefore, a bidding strategy involves planning across current and future auctions. European power market leader, Nord Pool, showed trades of 995 TWh of volume, with close to 60% of the volume traded using APIs [1]. Clearly, a bidding strategy that can optimize the cost of energy brokers even by a small amount would significantly improve profits and make the system more efficient.

In this paper, we first characterize Nash equilibria for single-buyer single-seller (SBSS) multi-unit (two identical and indivisible units) k-double auctions, where both buyer and seller follow scaling-based bidding strategies. Characterizing equilibrium becomes tractable beyond two units; hence, we design an intelligent agent (buyer) who can learn the equilibrium strategy. Motivated by the recent success stories of employing neural networks (NN) to solve game theoretical problems [14–17], we develop a DDPG-based bidding strategy DDPGBBS (technique derived from Reinforcement Learning (RL)) and perform validation experiments to show that it approximately achieves the theoretical equilibrium. We then extend DDPGBBS to work in general PDAs with no restriction on the number of participants or units traded in the auction. The strategy thus developed is then tested on a smart-grid ecosystem called PowerTAC, which is an efficient simulation of the real-world smart-grids; primarily, it simulates PDA for energy trading in the wholesale market [11]. We show that Extended DDPGBBS consistently outperforms benchmark and state-of-the-art PowerTAC bidding strategies. To the best of our knowledge, we are the first to utilize the policy gradient based RL algorithm in PowerTAC [4, 5, 8, 9, 12, 18–20, 22, 23], enabling us to work with PowerTAC’s continuous state and action space more effectively.

2 DDPGBBS BIDDING STRATEGY
Notation: Let us assume that the true types (valuations) of buyer B and seller S are θB and θS, respectively. Both B and S place two bids/asks in the auction by following scale-based bidding strategies bB and bS, respectively. The bB[bS] is defined as a strategy in which the buyer[seller] places two bids[asks] bB[1] = αB1θB and bB[2] = αB2θB [bS[1] = αS1θS and bS[2] = αS2θS]. Here, αB1 and αB2 (αS1 and αS2) are the scale factors by which B [S] scales its true type.

The below theorem presents our main results. We refer the reader to the extended version of our paper for the complete analysis [3].

Theorem 2.1. For an SBSS two-unit k-double auction with k = 0.5, where θB ~ U[θB, hB] and θS ~ U[θS, hS], respectively; when they deploy scale-based bidding strategies bB and bS, we get a system of equations, solving which results in a unique set of scale factors for the buyer and the seller that constitute a Bayesian Nash equilibrium.
DDPGBBS–Learning Equilibrium Bidding Strategies: Due to analytical intractability beyond SBS by two-units auctions for scale-based strategies, we train a DDPG-based strategy to learn equilibrium scale factors for the buyer. First, we study if DDPGBBS is able to learn the known equilibrium as follows: The state-space \( S \) consists of \textit{quantity to buy} \( q \), where \( q \in Q = \{0, 1, 2\} \); and buyer’s \textit{true type} \( \theta \in [0, 1] \). The actions are the buyer’s scale-factors \( a_{B1}, a_{B2} \in [0, 1] \). Buyer receives reward \( r = 0 \) if no market-clearing happens, else it receives reward \( r = -cp \cdot q + cq \) (where \( cp \) and \( cq \) are clearing price and buyer’s clearing quantity, respectively). An optimal strategy would be one that maximizes the expected reward. A state \( s \) transitions to the next state, where \textit{quantity to buy} \( q' \) is the remaining quantity \( q' \in Q = \{0, 1, 2\} \) after auction-clearing in state \( s \), while \( \theta \) remains the same. We consider a single-shot auction; thus, the episode terminates after a single step, and the buyer receives a terminal reward \( r = -q' + \theta \). DDPGBBS follows a similar configuration as described in [13]. However, due to the smaller state and action spaces, the NN used for the actor and critic had two hidden layers with only 40 and 30 units, respectively.

We perform controlled experiments to validate that DDPGBBS empirically follows the obtained theoretical equilibrium for each case by following the same assumptions used for theoretical analysis. The empirical result obtained using DDPGBBS is within 12.2\% of theoretical \( \alpha_{B1} \) for \{\( \alpha_{B1} = \alpha_{B2}, \alpha_{S1} = \alpha_{S2} \)\} case. Similarly, for other cases, too, empirical results are reasonably close to the theoretical results (refer Section 5 in [3]). Additionally, DDPGBBS showed low variances for all the scale factors, reinforcing its stability.

Extended DDPGBBS–Bidding Strategy for Smart-grids: As PDA allows multiple auction instances for a delivery slot, our DDPGBBS needs to be updated accordingly; thus, we propose Extended DDPGBBS to be helpful in general PDAs. We use PowerTAC’s wholesale market PDA to test the Extended DDPGBBS and use the ZI strategy to train it (The ZI strategy follows a randomized approach to bid in a PDA by sampling a price from a uniform distribution between the minimum bid price and maximum bid price). Below are the modifications incorporated in Extended DDPGBBS.

The state-space \( S \) includes an additional parameter \textit{proximity} \( p \in P = \{0, 1, 2, ..., 24\} \) and expands the domain of \textit{quantity to buy} \( q \). We consider the buyer’s \textit{true type} \( \theta \in R \) as the average \textit{unit balancing price} for \textit{buying} from PowerTAC’s balancing market in a game. During the game, Extended DDPGBBS outputs two scale-factors \( \{\alpha_{B1}, \alpha_{B2} \in [0, 1]\} \) which get multiplied with \( \theta \) to form the two bids in the auction, while the required bidding quantity is equally distributed into these two bids. The reward function remains the same, except, in a terminal state, it receives reward \( r = -q' + \theta \), where \( q' \) is the remaining quantity in the terminal state \( T \). After each auction, state transition occurs, \textit{Proximity} changes from \( p \) to \( p-1 \), \textit{quantity to buy} \( q \) becomes the remaining quantity after auction clearing, and \textit{buyer’s true type} \( \theta \) remains the same. The episode ends in \( T \), either when \( p = 0 \) or when \( q = 0 \).

Extended DDPGBBS is trained offline by collecting experiences in the replay buffer, using the PowerTAC PDA simulator. To collect experiences, we run two sets of experiments. In the first [second] set, Extended DDPGBBS competes against one [three] ZI broker[s] in two-player [four-player] games. In each set, we distribute hourly demand equally between all the competing brokers to make each broker participate equally in the wholesale market PDA. It updates the replay buffer after each auction instance in the game. After the execution of both sets is completed, we update Extended DDPGBBS using the combined replay buffer of both sets by following the standard DDPG update procedure. We train Extended DDPGBBS against ZI brokers as they do not follow any particular bidding pattern, and thus Extended DDPGBBS gets to see a wide range of states in the state-space, which improves its learning. Note that, unlike some previous PowerTAC brokers, Extended DDPGBBS does not incorporate any additional heuristics in its bidding strategy.

**Experiments and Results:** We benchmark the performance of Extended DDPGBBS against baseline and the state-of-the-art strategies using isolated PowerTAC wholesale market PDA. We perform two batches of experiments; the first batch of experiments is divided into four sets. In each of these four sets, we play ten two-player games between Extended DDPGBBS and one of the broker from the set \{SPOT [4, 5], VV [8], ZIP [6], ZI\}. Similarly, in the second batch, we play ten five-player games having all the available brokers in the game. In the first batch of experiments, we compare the average unit clearing price of the opponent in each set with respect to the Extended DDPGBBS’s clearing price; a value greater than 1 indicates that the opponent in that set had a higher average clearing price than Extended DDPGBBS after playing ten games. As shown in Table 1, Extended DDPGBBS outperforms all the other bidding strategies consistently by at least 9.9\% in two-player games while achieving a 33.76\% improvement against ZI. In the second batch, we compare each broker’s average unit clearing price across ten games in five-player games. Here too, as shown in Figure 1, Extended DDPGBBS consistently outperforms all the other brokers by at least 21.42\% (against second-best VV), while achieving almost 46\% improvement against other brokers.

### 3 CONCLUSION

We presented a Nash equilibrium analysis and showed that DDPGBBS-based bidding strategy, DDPGBBS, approximately achieves theoretical equilibrium. DDPGBBS can adapt to the increasing number of participating players and items in the real-world PDAs. We examined the efficacy of our novel bidding strategies against baseline and the state-of-the-art bidding strategy of PowerTAC PDA, where it consistently outperforms some of the best bidding strategies.
REFERENCES


