Multi-Agent Learning in Repeated Double-side Auctions for Peer-to-peer Energy Trading

Zibo Zhao
Google
Mountain View, CA
zibozhao@google.com

Andrew L. Liu
School of Industrial Engineering, Purdue University
West Lafayette, IN 47906
andrewliu@purdue.edu

Abstract

Distributed energy resources (DERs), such as rooftop solar panels, are growing rapidly and are reshaping power systems. To promote DERs, feed-in-tariff is usually adopted by utilities to pay DER owners certain fixed rates for supplying energy to the grid. Such a non-market based approach may increase electricity rates and create inefficiency. An alternative is a market based approach; i.e., consumers and DER owners trade energy in a peer-to-peer (P2P) market, in which electricity prices are determined by real-time market supply and demand. A prevailing approach to realize a P2P marketplace is through double-side auctions. However, the auction complexity in an energy market and the participants' bounded rationality may invalidate many well-established results in auction theory and hence, cast difficulties for market design and implementation. To address such issues, we propose an automated bidding framework based on multi-agent, multi-armed bandit learning through repeated auctions, which is aimed to minimize each bidder's cumulative regret. Numerical results suggest the potential convergence of such a multi-agent learning game to a steady-state. We also apply the framework to three different auction designs (including uniform-price versus Vickrey-type auctions) for a P2P market to study the impacts of the different designs on market outcomes.

1. Introduction

Distributed energy resources (DERs), broadly termed to include distributed generation, storage and demand response resources, are a vital part of a smart energy grid, as such resources are clean and sustainable, compared to fossil-fueled power plants, and can improve system reliability and resilience with their proximity to load. Many difficulties exist, however, in realizing the perceived benefits of DERs. Such difficulties are largely contributed by the physical characteristics of an electric energy grid, and the complexities of designing a market upon such a grid. As stated in [1], there are two layers in utilizing DER resources - the physical layer, and the virtual layer.

At the physical layer, the supply and demand of electric energy needs to be balanced at all time, and electric currents travel based on Kirchhoff's laws through transmission/distribution networks. Such attributes call for a centralized dispatch operation, such as that organized by independent system operators in the U.S. wholesale markets. However, the distributed feature and the potentially very large number of DERs make such centralized dispatch extremely challenging. In addition, without access to a high voltage transformer (which most DERs do not have access to), energy generated by DERs cannot travel far, due to significant energy losses at low voltage level.

The virtual layer mainly concerns the market mechanisms (especially pricing mechanisms) and information exchange among market participants/consumers. While supply bidding (or double-side auction with demand bidding) and uniform-pricing (aka locational marginal pricing, or LMP) have been the norm for the U.S. wholesale energy markets [2], the mechanisms to integrate DERs are much less clear. If DERs are to participate in a wholesale market, (besides the voltage/transmission issues), they need aggegators to pool a large quantity of DERs together to be eligible to bid into a wholesale markets. In doing so, DER owners will have to relinquish control of their resources to the aggregators. Instead of directly participating to a wholesale markets, many have proposed to organize a distribution market, similar to the wholesale counterpart, to dispatch local DERs, and to clear the market through the so-called distribution LMP (or DLMP) (e.g., [3, 4]. The DLMP pricing mechanism, in our view, would encounter significant issues in a DER market. While wholesale markets are dominated by fossil-fueled plants with varying marginal costs (which can then form an upward sloping aggregated supply curve), a local DER market is expected to be dominated by small solar and wind

URI: https://hdl.handle.net/10125/70995
978-0-9981331-4-0

resources, which have practically zero marginal costs, and demand response resources, whose marginal benefits are difficult to quantify.[1] With energy storage, the issues would be worse since storage resources can arbitrate prices over time. Hence, we expect a renewable-dominated local energy markets would yield zero DLMPs most of the time (unless there are persistent transmission constraints).

An alternative pricing mechanism to DLMP is through a double-side auction in a peer-to-peer (P2P) trading market, in which local energy producers (referred to as prosumers) and consumers can freely bid/ask their price and quantity, and energy prices are determined at the intersection of the supply and demand curves. This is the focus of this paper. While auction designs have been well studied in the field of economics and game theory [5, 6, 7, 8], several special features of a peer-to-peer (P2P) energy market require special attention. To name a few, a P2P energy market inherently involves repeated auctions and exogenous uncertainties (e.g., wind/solar availability), making the analysis of market participants' bidding/asking strategies much more difficult. In addition, market participants are likely to have bounded rationality in the sense that they do not know their own valuation of energy production and consumption. Furthermore, their (implicit) valuations are likely dependent, such as in a hot summer day, most buyers would value high of energy consumption for air conditioning. All these features would nullify the assumptions in classic auction theory as they usually require that agents' valuation to be private and independent [9].

Despite the theoretical difficulties, P2P-based energy market for DERs and prosumers have attracted increasing attention, as summarized by several recent review papers [10, 11, 12, 1]. Among the reviewed works, several of them directly study the design and implementation of a double-side auction, including (but are not limited to) [13, 14, 15]. Our proposed framework differs significantly from existing works in the literature in the following aspects. (i) Acknowledging bidding agents' bounded rationality, we do not assume that the agents know their own valuation of energy consumption (such as the so-called utility function). Each agent only needs to know their own past bids and the corresponding payoffs. (ii) We explicitly consider repeated auctions[2] (as opposed to a single-shot auction), and define the metric – cumulative regret – to gauge each agent's bidding performance. (iii) The bidding strategies can be fully automated and implemented through the so-called energy management systems (or even a smart meter).

More specifically, our framework is based on a multi-agent, multi-armed bandit learning approach. Consider a double-side auction where supply and demand bids are submitted in each time period h (e.g., hourly) in each day d. Within each h, as a starting point, we assume that market participants only choose a price to ask/bid, not quantities of energy. We further discretize per-unit price bids (i.e., ¢/KWh) into K possible choices. When each agent decides which price to bid/ask, it is similar to choosing one slot machine, out of K such machines, to pull the arm. In this case, the agents are uncertain if they will win (bids cleared) or lose (bids not cleared), and in the case of winning, how much the payoff would be. This is similar to the classic multi-armed bandit (MAB) learning problem which has been well studied in the broad computer science literature, such as [16, 17, 18, 19]. A key difference here, however, is that each agent's probability of winning and their payoff distributions (of each arm) depend on how other market participants bid/ask, and a MAB-game is formed when all agents apply bandit learning for deciding their bid/ask with incomplete information feedback.

The MAB-game differs from a pure agent-based simulation approach as certain theoretical results, such as convergence to a steady-state, can be obtained under certain conditions, as shown in [20]. Such a framework have been applied to integrate price-responsive demand response into a wholesale energy market with transmission constraints, and extended the theoretical results to show that each agent's cumulative regret will converge to zero [21, 22, 23]. In this paper, we show the details of how to apply the MAB-game framework in a double-side auction situation, which is by no means trivial. While establishing theoretical results is not the main goal of this paper, our numerical simulations do suggest of convergence to steady-state of the multi-agent MAB-game. In addition, we demonstrate the versatility of such a framework through studying three specific auction designs: a replicate of the wholesale market's uniform-price auction, a variant of Vickrey double-side auction [6], and maximum volume matching auction (which is pay-as-bid/receive-as-ask) [7]. Based on the simulations, from market participants' perspective, the uniform-price auction outperforms the other two as it can offer higher clear quantities, total social welfare and total normalized reward.

The rest of the paper is structured as follows. In Section 2, we describe in details of how market

[1] Or if there are combined heat and power plants in a local market that use natural gas, biomass or solid waste, they will be on the margin all the time, who then may choose not to participate in a local market at all.

[2] It is well-known in the economics and game theory literature that more diverse strategies can emerge, such as tacit collusion, in a repeated game than in a static game.

participants bid/ask through bandit learning in a double-side auction. In Section 3, three double-side auction mechanisms are presented for P2P energy trading market. Numerical simulations are presented in Section 4 by comparing learning results in three different auction mechanisms. Section 5 concludes the paper and identifies potential future research directions.

2. Learning through Multi-agent MAB Games

Without a P2P energy market, prosumers can only sell their extra energy to the utility or distribution system operator (DSO) at some pre-defined rate (such as through a feed-in tariff, denoted as FIT). Similarly, consumers can only buy energy from the utility under some pre-approved rates. In this work, we consider time-of-use (TOU) pricing, widely applied by utilities in the U.S., for customers buying energy, i.e a fixed rate for each time period (e.g. hourly). While in a bilateral P2P marketplace, consumers and prosumers can trade with each other at rates accepted by both buy-side and sell-side. Intuitively, a marketplace is desired by both sides if it can provide agents with some rate higher than FIT for sellers, and lower than TOU for buyers. Otherwise, agents can simply sign contract with the utility to buy/sell at TOU/FIT.

To incentivize the growth of DERs, a double-side auction can be organized for clearing bids/asks from market participants in each time period. Agents need to decide their unit price and quantities of energy for submitting bids/asks to the auctions. In this work we assume that with some smart devices using historical and weather data, agents can accurately forecast how much energy themselves will consume or generate in very near future (e.g. in one hour), and thus quantities can be easily decided for the auctions. It is the bid/ask price that is difficult to decide. To address the issue, we propose a MAB-game learning approach for a multi-agent system in which bidding/asking prices of agents are automatically chosen by bandit learning algorithms (and hence, can be easily implemented by control automation devices without human intervention). For illustration purposes, the herein presented formulas concern a single trading-period h (e.g. 1 hour) across days. We consider a set of agents $\mathcal{A} = \mathcal{A}_b \cup \mathcal{A}_s$, where $\mathcal{A}_b$ and $\mathcal{A}_s$ are the sets of buyers and sellers, respectively. Further, we let P_{FIT} and P_{TOU} denote the FIT and TOU rate in ¢/KWh, respectively, and we only consider the situation where the FIT is lower than the TOU rate, i.e. $P_{FIT} < P_{TOU}$.

2.1. Discrete Price Arms

The majority of DERs are solar and wind resources, and thus we consider their generation marginal costs as zero despite of fixed installment and maintenance fees. Therefore, any rate higher than FIT would be attractive to DER owners. Similarly, energy buyers desire for any rate lower than TOU rate. Therefore, any rate (in ¢/KWh) in the range $[P_{FIT}, P_{TOU}]$ would profit both energy buyers and sellers, and any reasonable agent $i \in \mathcal{A}$ has a bidding/asking price space $\mathcal{P}_i \in \mathcal{Z}^{\geq 0}$ which contains both P_{FIT} and P_{TOU}.

Herein, each discrete unit price in space $\mathcal{P}_i$ is a price arm that can be picked up for the agent's bid/ask. How to choose a price arm is complicated due to the dynamics of auctions. For each individual agent, it prefers a lower/higher auction clear price if it is a buyer/seller. However, it is not necessary that an agent's bidding/asking price is the auction clear price which depends on the collection of bids and asks. Since agents are not bidding/asking based on their implicit valuations (which are not known by agents), under some auction designs, like uniform price double auction, some agents may take chance by bidding/asking some extreme high/low unit price to make their bids/asks more likely to be accepted by the auction while enjoy the more profitable clear price. In Section 2.3, we will discuss the performance bound (i.e regret bound) of picking up price arms in the auction games by bandit learning for agents.

2.2. Rewards

The reward (or payoff) each agent receives in the auction represents the normalized level of the actual sent/received payment, Λ_i, between the lower and upper benchmarks by P_{TOU} and P_{FIT} which are denoted by $\underline{\Lambda_i}$ and $\overline{\Lambda_i}$, respectively. Herein, we let q_i denote the demand/supply of agent i, and q_i is negative for a buyer and positive for a seller, i.e. $q_i < 0|_{i\in\mathcal{A}_b}$ and $q_i > 0|_{i\in\mathcal{A}_s}$. For a buyer agent, the lower and upper benchmarks refer to buying all of q_i at P_{TOU} and P_{FIT}, respectively. In the opposite, a seller agent has its lower and upper benchmarks with selling q_i at P_{FIT} and P_{TOU}. Therefore, we have

$$\underline{\Lambda_i} = q_i \cdot [P_{TOU} \cdot \mathbb{1}_{\{i\in\mathcal{A}_b\}} + P_{FIT} \cdot \mathbb{1}_{\{i\in\mathcal{A}_s\}}], \quad (1)$$

$$\overline{\Lambda_i} = q_i \cdot [P_{FIT} \cdot \mathbb{1}_{\{i\in\mathcal{A}_b\}} + P_{TOU} \cdot \mathbb{1}_{\{i\in\mathcal{A}_s\}}]. \quad (2)$$

The actual sent/received payment of each agent $\forall i \in \mathcal{A}$ consists of two parts for trading in the auction and with the utility, which are denoted by Λ_i^{au} and Λ_i^{ut},

respectively. Thus, we have

$$\Lambda_i = \Lambda_i^{au} + \Lambda_i^{ut}. \tag{3}$$

With attending the auction, market participants send/receive payments based on the clear result. Specifically, each agent's sent/received payment in the auction is calculated according to its clear price, p_i^{au}, and clear quantity, q_i^{au}, as below

$$\Lambda_i^{au} = p_i^{au} \cdot q_i^{au}. \tag{4}$$

In auctions like uniform price double auction, all agents have the same clear price. While in the maximum volume matching auction [7], the agents may have different clear price since they pay/receive at their bid/ask price.

However, it is not necessary that all agents are buying/selling in the P2P market since some agents' bids/asks may not be (fully) cleared by the market. In this case, for not wasting the (renewable) energy from DERs, prosumers are allowed to sell the unclear energy to the utility at P_{FIT}. Also, consumers always can buy their demand not satisfied by the P2P market from the utility at P_{TOU}. Therefore, the sent/received payment to/from the utility for agent i is as below

$$\Lambda_i^{ut} = p_i^{ut} \cdot q_i^{ut}, \tag{5}$$

where $p_i^{ut} = P_{FIT}$ if $i \in \mathcal{A}_s$ and $p_i^{ut} = P_{TOU}$ if $i \in \mathcal{A}_b$, and q_i^{ut} denotes the unclear energy quantity.

Then we have $q_i = q_i^{au} + q_i^{ut}$. When the agent's auction clear price $p_i^{au} \in [P_{FIT}, P_{TOU}]$, we have $\Lambda_i \in [\underline{\Lambda_i}, \overline{\Lambda_i}]$ and thus we have the normalized reward $\pi_i \in [0, 1]$ calculated as below

$$\pi_i = (\Lambda_i - \underline{\Lambda_i})/(\overline{\Lambda_i} - \underline{\Lambda_i}). \tag{6}$$

In Eq. (6), we can see for $p_i^{au} = P_{FIT}$, a buyer agent has $\pi_i = 1$ while a seller agent has $\pi_i = 0$, and for $p_i^{au} = P_{TOU}$ we have the opposite values. However, in Section 2.1 we mentioned that the agent's bidding/asking price space $\mathcal{P}_i$ contains P_{FIT} and P_{TOU}, and thus the agent may bid/ask some price outside the range $[P_{FIT}, P_{TOU}]$. Though it is counter-intuitive, the auction clear price p_i^{au} could be outside $[P_{FIT}, P_{TOU}]$, even in the uniform-price double auction if a significant population are doing so. In the case $p_i^{au} < P_{FIT}$, we consider $\pi_i = 1|_{i \in \mathcal{A}_b}$ and $\pi_i = 0|_{i \in \mathcal{A}_s}$; for $p_i^{au} > P_{TOU}$, $\pi_i = 0|_{i \in \mathcal{A}_b}$ and $\pi_i = 1|_{i \in \mathcal{A}_s}$. Combined with Eq. (6), we have $\pi_i =$

$$\begin{cases} 1 \cdot \mathbb{1}_{\{i \in \mathcal{A}_b\}} + 0 \cdot \mathbb{1}_{\{i \in \mathcal{A}_s\}}, & \text{for } p_i^{au} < P_{FIT} \\ (\Lambda_i - \underline{\Lambda_i})/(\overline{\Lambda_i} - \underline{\Lambda_i}), & \text{for } P_{FIT} \le p_i^{au} \le P_{TOU} \\ 0 \cdot \mathbb{1}_{\{i \in \mathcal{A}_b\}} + 1 \cdot \mathbb{1}_{\{i \in \mathcal{A}_s\}}, & \text{for } p_i^{au} > P_{TOU}, \end{cases} \tag{7}$$

where $\underline{\Lambda_i}$, $\overline{\Lambda_i}$, and Λ_i can be achieved by Eq. (1), (2), and (3), respectively.

2.3. Pricing by Bandit Learning

As in Eq. (7), we can see the reward π_i of each agent highly depends on its clear price in auction which further depends on its bid/ask and the collection of other agents' bids/asks. The dynamic auction games result in nonstationary clear prices, which makes the bidding/asking decision-making difficult for agents. In regular game theory literature, the standard equilibrium concept for dynamic games of incomplete information is Perfect Bayesian Nash equilibrium (PBNE) [24, 25]. In a PBNE, the collection of each agent's action profile maps the entire history of the games to each agent's feasible set of actions, under the assumption that each agent maintains their beliefs of other competitors' distribution of action space based on the Bayes' updating rule. For a large population, the assumption requirement is impractical and implausible for small-scale (in terms of computation power) agents in P2P energy trading auctions. This is where MAB-game comes in. Instead of tracking their competitors' tremendous states, agents only need to look at their own history in repeated games. A recent breakthrough on MAB-game in [20] has provided us with the theoretical foundations in studying the auction games with a large population in this work. A key point in MAB-game with many agents is that as every agent conducts its own stochastic *no-regret* bandit learning independently in repeated games, the finite system will approximately converge to the unique *mean field steady state* (MFSS) of the infinite population system. The population profile (i.e. the proportion of population on each arm) is stationary in the MFSS, and the approximation gets better as the finite population increase. Under the stationary population profile, efficient outcomes will be achieved since each individual agent can solve its MAB problem with stationary reward distributions as in classic MAB problem settings.

We let $\boldsymbol{f}$ denote the energy quantities' stationary population profile of the agent set $\mathcal{A}$, where $f(k)$ represent the distribution of buying and selling energy quantities on price arm k. With stationary population profile $\boldsymbol{f}$, each agent has its underlying optimal bid/ask price arms whose associated clear price results in the optimal reward as below

$$\pi_i^*(\boldsymbol{f}) = \max_{k \in \mathcal{P}_i} \mathbb{E}[\pi_i(\boldsymbol{f}, k)], \tag{8}$$

where $\pi_i(\boldsymbol{f}, k)$ denotes the reward of agent i for picking up price arm k under population profile $\boldsymbol{f}$.

Suppose that for the trading-period h across D days (i.e. D rounds in our context), agent i uses a policy σ which is an algorithm picking up the next price arm based on its learning history. The history is only about the agent's own sequence of played price arms and corresponding observed rewards, which largely reduces the knowledge dimension that the agent has to maintain. Though the underling optimal reward $\pi_i^*(\boldsymbol{f})$ is unknown to the agent, the policy σ enables the agent to learn about the distributions of rewards for each price arm. Let $\Gamma_\sigma(D,k)$ be the number of times price arm k has been picked up by the policy σ during all the D rounds. Then for agent i, we define its cumulative regret under the policy σ for every D rounds as below

$$\Delta_\sigma = \pi_i^*(\boldsymbol{f}) \cdot D - \sum_{k \in \mathcal{P}_i} \mathbb{E}[\pi_i(\boldsymbol{f}, k) \cdot \Gamma_\sigma(D,k)]. \quad (9)$$

The regret Δ_σ in Eq. (9) is the expected loss due to the fact that the policy does not necessarily always pick up the optimal price arm under the stationary population profile which is unknown to the agent. The policy σ is a *no-regret* bandit learning policy if the regret in Eq. (9) satisfies:

$$\frac{1}{D}\Delta_\sigma < R(D,K), \quad (10)$$

for some $o(1)$ function R in terms of D; where K is the cardinality of $\mathcal{P}_i$, i.e. $|\mathcal{P}_i| = K$. Then $R(D,K)$ gives an upper bound to the average regret under the policy σ. For the bandit learning algorithms based on UCB [17], such as UCB1, UCB-tuned and UCB2, we have logarithmic regret bounds that are $o(1)$ in terms of total rounds D: $R(D,K) = \alpha(K) \cdot \frac{1}{D} ln(D)$. Therefore, as the auction games go on, the agent's average regret goes to 0.

3. Double Auction Designs

In this section, we first define the individual monetary utility, corresponding total social welfare, and auctioneer's profit with a P2P energy market auction. Then we discuss about three different double-side auction designs that can be applied for the market clear: the uniform-price auction, a variant of Vickrey double-side auction [6], and the maximum volume matching auction [7].

3.1. Social Welfare and Auctioneer's Profit

As mentioned above, agents are rarely aware of their private valuation of energy production and consumption. To define agents' individual monetary utility, we consider it as profit for energy sellers and costs reduction for buyers with participating the P2P market. Since for renewable DER owners, the marginal cost is almost zero, the total profit of energy seller $i \in \mathcal{A}_s$ is as below

$$u_i|_{i \in \mathcal{A}_s} = p_i^{au} \cdot q_i^{au} + P_{FIT} \cdot q_i^{ut}, \quad (11)$$

which has the same value as Λ_i in Eq. (3). For consumers, they have to pay at P_{TOU} without the P2P market, thus we have the cost reduction as

$$u_i|_{i \in \mathcal{A}_b} = (P_{TOU} - p_i^{au}) \cdot |q_i^{au}|. \quad (12)$$

In spite of the auctioneer's profit, the total social welfare of all agents, denoted by $U_\mathcal{A}$, is simply the aggregation of all agents' utility, i.e. $U_\mathcal{A} = \sum_{i \in \mathcal{A}} u_i$. For the auctioneer (which can be played by the utility or DSO), the total auction trading surplus it earns is the sum of bid-ask price difference for each energy unit traded in the auction, which is calculated as below

$$U_\mathcal{M} = \sum_{i \in \mathcal{A}_b} (p_i^{au} \cdot |q_i^{au}|) - \sum_{i \in \mathcal{A}_s} (p_i^{au} \cdot q_i^{au}), \quad (13)$$

where $U_\mathcal{M}$ denotes the auctioneer's profit.

3.2. Uniform-Price Double Auction

If price is plotted as a function of aggregate energy quantity following the convention in economics, then the energy demand and supply curves slope downward and upward, respectively, as shown in Fig. 1. Graphically, the intersection (P^*, Q^*) of the supply and demand curves clears the market at which the quantity demanded is equal to the quantity supplied. The price P^* is the *equilibrium price*, and the corresponding energy quantity is the *equilibrium quantity*. As such, all agents pay/receive at the uniform price P^*, and the quantity Q^* in total is traded in the auction. Then the rest supply $Q_s - Q^*$ is sold to the utility at P_{FIT}, in which Q_s denotes the total energy supplied by DERs, i.e. $Q_s = \sum_{i \in \mathcal{A}_s} q_i$. Also, the unsatisfied demand is purchased from the utility at P_{TOU}. Therefore, in Fig. 1, the shadow area in light purple represents the total social welfare $U_\mathcal{A}$, i.e.

$$U_\mathcal{A} = P_{TOU} \cdot Q^* + P_{FIT} \cdot (Q_s - Q^*). \quad (14)$$

Since $p_i^{au} = P^*$ for all agents $i \in \mathcal{A}$, and both $\sum_{i \in \mathcal{A}_b} |q_i^{au}|$ and $\sum_{i \in \mathcal{A}_s} q_i^{au}$ are equal to Q^*, by Eq. (13) the auctioneer earns zero profit in the auction, i.e. $U_\mathcal{M} = 0$.

3.3. Vickrey Variant Double Auction

Instead of paying/receiving at the uniform *equilibrium price*, we consider a Vickrey-like auction,

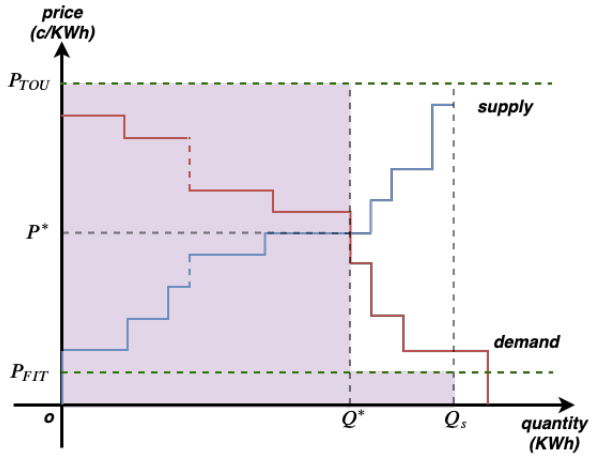

Figure 1: A uniform-price double auction market.

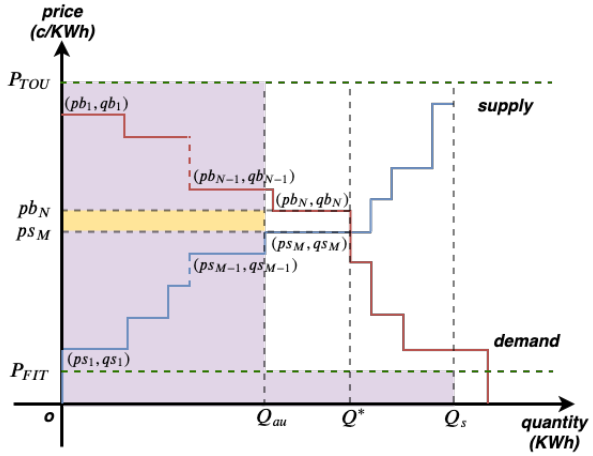

Figure 2: A Vickrey-like double auction market (Case I).

similar to that in [26]. The mechanism works as follows: similar to the uniform-price auction, all bids/asks are sorted down/up by bid/ask price, and we can have stair-wise demand/supply curves as shown in Fig. 2, in which each stair represents the bid/ask pair of price and quantity. We use subindex n to denote buyers, and m to denote sellers. Without loss of generality, we can assume that buyers' bid prices are $pb_1 > pb_2 > \cdots > pb_n > pb_{n+1} > \cdots$;[3] similarly, the sellers' ask prices are $ps_1 < ps_2 < \cdots ps_m < ps_{m+1} < \cdots$. At the critical intersection point (P^*, Q^*) where the aggregate demand and supply meet, there are N buyers and M sellers that are cleared. We consider the two possibilities at the clearing: Case I (as shown in Fig. 2):

$$pb_N \geq ps_M \geq pb_{N+1}, \tag{15}$$

$$\sum_{m=1}^{M-1} qs_m \leq \sum_{n=1}^{N} qb_n \leq \sum_{m=1}^{M} qs_m, \tag{16}$$

and Case II:

$$ps_{M+1} \geq pb_N \geq ps_M, \tag{17}$$

$$\sum_{n=1}^{N-1} qb_n \leq \sum_{m=1}^{M} qs_n \leq \sum_{n=1}^{N} qb_m. \tag{18}$$

Here we only describe the clearing mechanism for Case I, as Case II is similar.

Rule 1 If $\sum_{n=1}^{N-1} qb_n \geq \sum_{m=1}^{M-1} qs_m$, there is overdemand. All the asks with $m < M$ sell all their supply qs_m at price ps_M; all the asks with $m \geq M$ sell their supply at P_{FIT} to the utility. All the bids with $n < N$ buy at pb_N and each of them buys a volume equal to $qb_n - (\sum_{n=1}^{N-1} qb_n - \sum_{m=1}^{M-1} qs_m)/(N-1)$; all the unsuccessful bids buy at P_{TOU} from the utility.

Rule 2 If $\sum_{n=1}^{N-1} qb_n \leq \sum_{m=1}^{M-1} qs_m$, there is oversupply. All the bids with $n < N$ buy all their demand qb_n at price pb_N; all the bids with $n \geq N$ buy their demand at P_{TOU} from the utility. All the asks with $m < M$ sell at ps_M and each of them sells a volume equal to $qs_m - (\sum_{m=1}^{M-1} qs_m - \sum_{n=1}^{N-1} qb_n)/(M-1)$; all the unsuccessful asks sell at P_{FIT} to the utility.

According to the clear rules, the total trade volume in the auction is

$$Q_{au} = min(\sum_{n=1}^{N-1} qb_n, \sum_{m=1}^{M-1} qs_m). \tag{19}$$

Then the total social welfare for all agents can be calculated as below (which is represented by the light purple area in Fig. 2)

$$U_{\mathcal{A}} = [(P_{TOU} - pb_N) + ps_M] \cdot Q_{au} + P_{FIT} \cdot (Q_s - Q_{au}). \tag{20}$$

The auctioneer's profit represented by the yellow shadow area in Fig. 2 is as below

$$U_{\mathcal{M}} = (pb_N - ps_M) \cdot Q_{au}. \tag{21}$$

3.4. Maximum Volume Matching Double Auction

Other than chasing socail welfare for agents or profit for acutionner, the auction design proposed in

[3]Note that the total number of bidders is irrelevant here; only the number of cleared buyers and sellers matter.

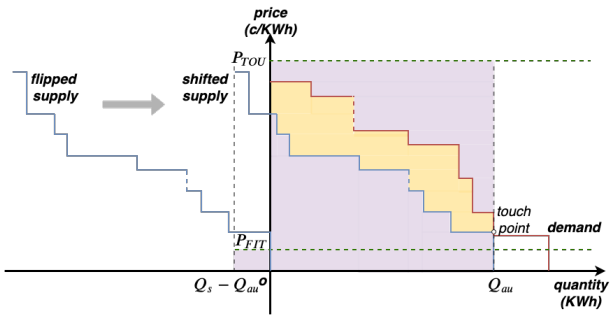

Figure 3: A maximum volume matching auction market.

[7] is for maximizing the traded volume given a set of bids and asks. The idea of market clear can be intuitively and graphically illustrated in Fig. 3. Suppose the demand/supply curves are based on the bids/asks shown in Fig 1. The supply curve is flipped horizontally and then shifted right towards the demand curve until the two curves touch. The distance, denoted by Q_{au}, that it can move is the minimal horizontal distance between the flipped supply curve and the demand curve which is exactly the maximum trading volume of the auction can be achieved. Then for the energy quantity 0 through Q_{au}, the corresponding bids (pb_n, qb_n) on the demand curve and asks (ps_m, qs_m) on the shifted supply curve are matched, and then successfully matched buyers/sellers pay/receive at their bid/ask price, respectively. Let $\mathcal{S}_b$ and $\mathcal{S}_a$ denote the set of successful bids and asks, respectively. The supply amount $Q_s - Q_{au}$ of the unsuccessful asks is sold to the utility at P_{FIT}, and also the unsatisfied demand is bought at P_{TOU}.

According to the clear mechanism, the total social welfare of all agents is as below (represented by the light purple shadow area in Fig. 3)

$$U_{\mathcal{A}} = \sum_{n\in\mathcal{S}_b}(P_{TOU} - pb_n)qb_n + \sum_{m\in\mathcal{S}_a} ps_m qs_m + P_{FIT}(Q_s - Q_{au}). \tag{22}$$

The auctioneer's profit is still the auction trading surplus (represented by the yellow shadow area in Fig. 3) as below:

$$U_{\mathcal{M}} = \sum_{n\in\mathcal{S}_b}(pb_n \cdot qb_n) - \sum_{m\in\mathcal{S}_a}(ps_m \cdot qs_m). \tag{23}$$

4. Numerical Simulations

In this section, we present the simulation results with distributed bandit learning corresponding to the three double-side auction designs for P2P energy trading as described in the previous section.

4.1. Input Data

4.1.1. Decision epochs and temporal resolution As a starting point, we do not consider time-linking constraints in our models, and each trading window is independent of others in a day. The simulations presented herein concern a single one-hour trading period for the peak hour 17:00 - 18:00 across 300 days, i.e. $D = 300$.

4.1.2. TOU/FIT and decision space We consider fixed TOU/FIT across days, and we let $P_{TOU} = 11$ ¢/KWh and $P_{FIT} = 5$ ¢/KWh. All agents has the same decision space $\mathcal{P}$ that contains all the discretized price arms through 0 ¢/KWh to 14 ¢/KWh, and thus P_{TOU}/P_{FIT} are included in $\mathcal{P}$.

4.1.3. Bandit learning algorithms for pricing For picking up price arms to bid/ask in the auctions, each agent $i \in \mathcal{A}$ uniformaly chooses its bandit learning algorithms among UCB1, UCB-tuned, UCB2, and $\epsilon - greedy$. Interested readers can refer to [17] for the details of the algorithms.

4.1.4. Consumers and energy demand to buy In the numerical test cases, we simulate 2000 distributed residential household consumers that participate in the auctions, i.e. $|\mathcal{A}_b| = 2000$. According to the Residential Residential Energy Consumption Survey (RECS) by U.S. Energy Information Administration (EIA) [27], a residential customer consumes about 30 KWh per day on average. Consider it is a peak hour, we naively let consumers repeatedly sample their energy demand quantities from a *Uniform* distribution $U(1.5, 2)$ in KWh, independently, for the hour across days, which is slightly higher than the average consumption level.

4.1.5. Prosumers and energy supply to sell On the sell-side, we also consider 2000 prosumers with DERs, i.e. $|\mathcal{A}_s| = 2000$. For the DERs, we only consider two renewable resources, solar and wind, for small-scale distributed agents in this work.

Table 1: Wind turbine models

Model	KW Rating
Energy Ball HEA V100 1.1m	0.5
Bergey BWC XL.1	1
True North Power Arrow 2m	1.23
Future Energy FE1048U 1.8m	1.5
Hummer 3.1m	2
Energy Ball HEA V200 1.98m	2.23
Southwest Windpower Skystream 3.7m	2.63
Westwind 3.7m	3.1

Due to the popularity of distributed residential solar panels (especially in western), we assume $4/5$ of the prosumers have solar-based distributed generation, and the other $1/5$ have wind-based. In the simulations, we use System Advisor Model (SAM) [28] developed by National Renewable Energy Laboratory (NREL) to model residential generation by solar and wind. The weather resource data for Arizona State by NREL is used for the simulations in SAM.

For the solar generation, we consider all panels have nameplate capacity as 2 KWdc with DC to AC ratio of 1.2 and inverter efficiency of 96%. For each distributed solar resource owner, the module type and array type have equal chance to be one of {*Standard, Premium, Thin Film*} and {*Fixed Open Rack, Fixed Root Mount, 1 Axis Tracking, 1 Axis Backtracking, 2 Axis Tracking*}, respectively. All other inputs are set as default in the *Photovoltaic PVWatts* simulations for distributed residential in SAM. More details about photovoltaic simulations can be found in [28, 29, 30].

For the simulations of distributed residential wind generation, each wind-based prosumer samples its turbine model uniformly from the 8 wind turbine models listed in Table 1, and the number of turbines owned by the prosumer is uniformly sampled among 1 through 4. All other inputs are set as default in the *Wind Residential* simulations in SAM. The turbines' specifications, such as wind power curves and turbine layout, can be found in [28, 30].

4.2. Numerical Results

The three different auction designs are simulated with the input data. We use UP, VV, and MV to denote uniform price auction, Vickrey variant auction, and maximum volume matching auction, respectively.

In Fig. 4, the clear quantity results of the auctions are presented, and we can see the results all have a trend of convergence. The counter-intuitive phenomenon is that in the later phase, UP is more likely to have a higher level of traded volume than MV which is designed to maximize traded volume. The reason is that with bandit learning, agents are updating their bids/asks dynamically, and thus the collective bids/asks schedules are not necessarily the same for different auctions. Besides the volume, we can see after a while of learning, UP's total clear quantity has smaller volatility than the other two auction designs. Therefore, in terms of auction clear quantity, UP outperforms VV and MV, and thus the auction design can let more renewable DERs be utilized.

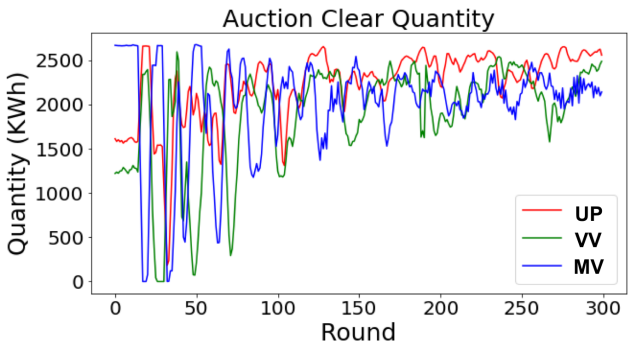

Figure 4: Total clear energy quantities (KWh) in the auctions.

Similar to the clear quantity, agents' total social welfare also display the convergence trend in the auctions, as shown in Fig. 5. Associated with more clear quantity, buyers and sellers in UP have higher social welfare (in $) than in the other two auctions in the later auctions. The performance of VV and MV are close to each other. Accordingly, for the total normalized reward, the results display very similar patterns as shown in Fig. 6.

Though UP outperforms the other two auctions for benefiting market participants and incentivizing DERs, it is not necessary that the auctioneer prefers it as well. As discussed in Section 3, the auctioneer has no profit in UP due to the zero trading surplus, which is validated by our simulations as shown in Fig. 7. According to the results, the auctioneer can achieve the most profit in MV, though the profit fluctuations of MV are much higher than VV's.

5. Conclusion

In this work, we propose a multi-agent MAB-game framework for market participants to (automatically) choose bid/ask prices in a P2P, double-side auction. The bandit learning approach allows each individual agent to make a decision only according to its own history, which is both privacy-preserving practical.

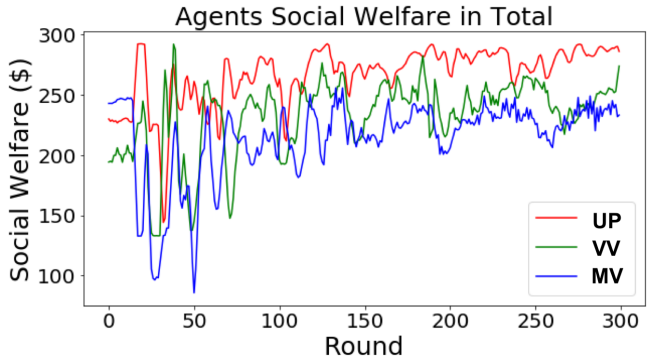

Figure 5: Total social welfare ($) of all buyers and sellers in the auctions.

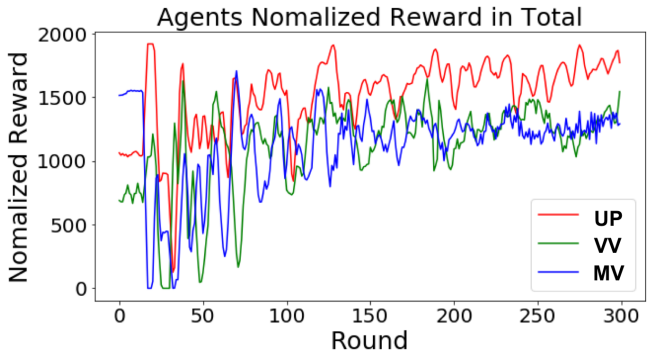

Figure 6: Total normalized reward of all buyers and sellers in the auctions.

We conduct simulations for the approach under three different auction designs, and the results suggest potential convergence of the cleared quantities, total social welfare and total normalized reward for agents. Moreover, the uniform-price double auction outperforms the other two in terms of market participants' benefits. For auctioneer, the maximum volume matching offers the highest profit.

While our work focuses on the virtual layer of utilizing DERs, as mentioned in the introduction section, it does not conflict with a local energy market operated by a distribution system operator (DSO). Indeed, a DSO is needed to maintain system feasibility of the cleared bids. This would be similar to the early days of the deregulated California wholesale energy market, which had a separate power exchange (for market clearing) and a system operator (for maintaining physical feasibility). Such a market structure has widely been considered as a main culprit for California electricity market's failure around year 2000 [31]. How to avoid such failure in a local P2P market, or in general,

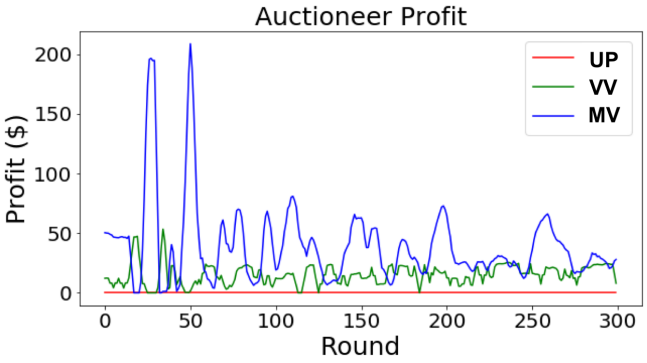

Figure 7: Auctioneer's profit ($) in the auctions.

how to maintain physical feasibility with the proposed market mechanism will be a immediate research task.

Another future research direction will be to investigate how blockchain technology can be utilized within the MAB-game framework, either at the virtual layer, or at the physical layer, to realize a fully decentralized energy market.

Acknowledgment

This research is partially supported by National Science Foundation grant ECCS-1509536 and CMMI-1832688, and the U.S. Department of Energy, Office of Electricity, under Award Number DE-OE0000921.

References

[1] W. Tushar, T. K. Saha, C. Yuen, D. Smith, and H. V. Poor, "Peer-to-peer trading in electricity networks: An overview," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3185 – 3200, 2020.

[2] P. Cramton, "Electricity market design," *Oxford Review of Economic Policy*, vol. 33, no. 4, pp. 589–612, 2017.

[3] R. Li, Q. Wu, and S. S. Oren, "Distribution locational marginal pricing for optimal electric vehicle charging management," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 203–211, 2013.

[4] F. Meng and B. H. Chowdhury, "Distribution lmp-based economic operation for future smart grid," in *2011 IEEE Power and Energy Conference at Illinois*, pp. 1–5, IEEE, 2011.

[5] D. Friedman, *The double auction market: institutions, theories, and evidence*. Routledge, 2018.

[6] P. Huang, A. Scheller-Wolf, and K. Sycara, "Design of a multi–unit double auction e–market," *Computational Intelligence*, vol. 18, no. 4, pp. 596–617, 2002.

[7] J. Niu and S. Parsons, "Maximizing matching in double-sided auctions," 2013. Available at arXiv.org: `https://arxiv.org/abs/1304.3135`, last revised: Feb 11, 2013.

[8] J. Nicolaisen, V. Petrov, and L. Tesfatsion, "Market power and efficiency in a computational electricity market with discriminatory double-auction pricing," *IEEE transactions on Evolutionary Computation*, vol. 5, no. 5, pp. 504–523, 2001.

[9] R. P. McAfee and J. McMillan, "Auctions and bidding," *Journal of economic literature*, vol. 25, no. 2, pp. 699–738, 1987.

[10] C. Zhang, J. Wu, C. Long, and M. Cheng, "Review of existing peer-to-peer energy trading projects," *Energy Procedia*, vol. 105, pp. 2563–2568, 2017.

[11] M. Khorasany, Y. Mishra, and G. Ledwich, "Market framework for local energy trading: a review of potential designs and market clearing approaches," *IET Generation, Transmission & Distribution*, vol. 12, no. 22, pp. 5899–5908, 2018.

[12] T. Sousa, T. Soares, P. Pinson, F. Moret, T. Baroche, and E. Sorin, "Peer-to-peer and community-based markets: A comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 104, pp. 367–378, 2019.

[13] W. Liu, D. Qi, and F. Wen, "Intraday residential demand response scheme based on peer-to-peer energy trading," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 1823–1835, 2019.

[14] K. Chen, J. Lin, and Y. Song, "Trading strategy optimization for a prosumer in continuous double auction-based peer-to-peer market: A prediction-integration model," *Applied energy*, vol. 242, pp. 1121–1133, 2019.

[15] J. Guerrero, A. C. Chapman, and G. Verbič, "Decentralized p2p energy trading under network constraints in a low-voltage network," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5163–5173, 2018.

[16] T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in applied mathematics*, vol. 6, no. 1, pp. 4–22, 1985.

[17] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2-3, pp. 235–256, 2002.

[18] W. B. Powell and I. O. Ryzhov, *Optimal learning*, vol. 841. John Wiley & Sons, 2012.

[19] A. Mahajan and D. Teneketzis, "Multi-armed bandit problems," in *Foundations and applications of sensor management*, pp. 121–151, Springer, 2008.

[20] R. Gummadi, R. Johari, S. Schmit, and J. Y. Yu, "Mean field analysis of multi-armed bandit games." Available at SSRN: `http://dx.doi.org/10.2139/ssrn.2045842`, last revised: August 11, 2016.

[21] Z. Zhao and A. L. Liu, "Intelligent demand response for electricity consumers: A multi-armed bandit game approach," in *2017 19th International Conference on Intelligent System Application to Power Systems (ISAP)*, pp. 1–6, IEEE, 2017.

[22] Z. Zhao, A. L. Liu, and Y. Chen, "Electricity demand response under real-time pricing: A multi-armed bandit game," in *Proceedings, APSIPA Annual Summit and Conference*, vol. 2018, pp. 12–15, 2018.

[23] Z. Zhao, "Decentralized price-driven demand response in smart energy grid." PhD dissertation, School of Industrial Engineering, Purdue University, 2019.

[24] D. Fudenberg and J. Tirole, "Game theory," *MIT press, Cambridge, Massachusetts*, vol. 393, no. 12, p. 80, 1991.

[25] D. Fudenberg, F. Drew, D. K. Levine, and D. K. Levine, *The theory of learning in games*, vol. 2. MIT press, 1998.

[26] P. Huang, A. Scheller-Wolf, and K. Sycara, "Design of a multi–unit double auction e–market," *Computational Intelligence*, vol. 18, no. 4, pp. 596–617, 2002.

[27] U.S. Energy Information Admisnistration (EIA), "Residential energy consumption survey (RECS)." available at `https://www.eia.gov/consumption/residential/data`.

[28] National Renewable Energy Lab (NREL), Golden, CO (United States), "System advisor model (SAM)." avaible at `https://sam.nrel.gov`.

[29] P. Gilman, "Sam photovoltaic model technical reference," tech. rep., National Renewable Energy Lab (NREL), Golden, CO (United States), 2015. avaible at `https://www.osti.gov/biblio/1215213`.

[30] N. Blair, A. P. Dobos, J. Freeman, T. Neises, M. Wagner, T. Ferguson, P. Gilman, and S. Janzou, "System advisor model, sam 2014.1.14: General description," tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2014. avaible at `https://www.osti.gov/biblio/1126294`.

[31] P. L. Joskow, "California's electricity crisis," *oxford review of Economic Policy*, vol. 17, no. 3, pp. 365–388, 2001.