# *Decentralized Control of Distributed Energy Resources via Multi-armed Bandit Games*

Zibo Zhao, Purdue University, West Lafayette, IN (917) 853-6199, zhao438@purdue.edu
Andrew L. Liu, Purdue University, West Lafayette, IN (765) 494-4763, andrewliu@purdue.edu

## Overview

Power systems worldwide are witnessing rapid growth of distributed energy resources (DERs), which are flexible and may be controlled by smart devices to help improve system efficiency and reliability, offset electricity price volatility, and promote renewable energy. However, due to the large number of local DERs, direct control by centralized authorities, such as system operators (e.g. ISOs) or utilities, are intractable and impractical. Besides, privacy issue can pose another hurdle. To directly address these difficulties, we propose a completely decentralized framework to integrate demand response and local energy generation into the grid under real-time pricing (RTP).

Under RTP, we assume that consumers' electricity bills (or profits of sending energy back to grid) are settled based on real-time (RT) electricity prices, as opposed to flat-rate or time-of-use prices. There has been a rich literature on how individuals should best respond to time-varying prices; however, much fewer works exist to study system-level impacts. Since all DERs receive the same price signals (such as day-ahead (DA) wholesale prices), naïve response would cause significant price volatility and system instability in real time. For example, if the DA price is high for an hour the next day, flexible demand would not use electricity in that hour; while prosumers (those who can generate electricity locally and send back to the grid) would all want to schedule their generation at the same hour. Such demand withdrawal, coupled with an influx of supply from DERs, would cause RT price to collapse and add significant pressure for the ISO to maintain supply-demand balance. In a decentralized setting, a game-theoretic approach is needed to avoid such a "herding" effect. Traditional equilibrium concepts would impose strong rationality assumptions on agents and are ill-equipped when the number of agents is large. Alternatively, we propose a framework within which agents have bounded rationality (i.e., they do not know other agents' utility functions and do not contemplate the impacts of their own actions to the game's outcome), and learn from their past actions. If each agent employs a regret-minimizing strategy, the overall system can be shown to approximate to a steady-state (termed as a mean field equilibrium) as time progresses. The approximation becomes exact when the number of agents goes to infinity, making this approach particularly scalable. Simulation results show that our approach can not only significantly reduce wholesale electricity price volatility, but also alleviate transmission congestion and increase total social surplus, when compared to the naïve-response approach.

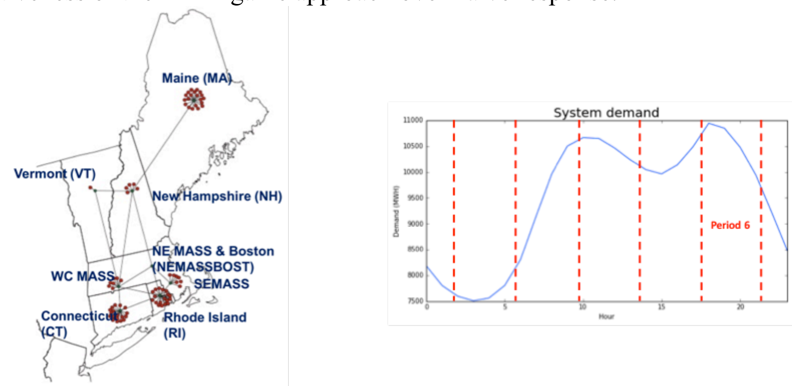
## Methods

We consider a two-settlement wholesale market where an ISO solves an economic dispatch (ED) problem to determine the hourly DA prices based on next-day demand forecast. Under naïve response, prosumers use the DA prices to determine the best actions for the next day. Real-time supply-demand imbalance is resolved by the ISO through solving another ED problem, which determines the RT prices. In our approach, termed as the multi-armed bandit game (or MAB-game) approach, prosumers treat each hour within a certain period next day (say, 24-hour, on-peak, off-peak, etc) as an arm of a mutli-armed slot machine, and pulling an arm means using electricity or generating energy at that particular hour. Once an arm is pulled, a payoff is realized, which is either the negative of the electricity cost or the profit of selling energy in that hour. Prosumers receive the information of the past RT prices, and would know, in hindsight, what would be best arm to pull. (The difference between the best possible payoff and the realized payoff is referred as regret.) As the same game is repeated daily, agents gradually learn which arm is the best to pull through exploring (trying as many arms as possible) and exploiting (keeping pulling the arm that gives the best cumulative payoff so far). An essence of the MAB game is that each arm's payoff depends on what the agents do collectively, as in the DER case in which prosumers' collective actions would affect RT prices. Under a generic and much simpler game setting, Gummadi et al.[1] show the existence and uniqueness of a steady-state (aka a mean-field equilibrium) of an infinite-agent MAB game when each agent uses a regret-minimizing strategy; in addition, a finite agent MAB game can uniformly converge to the mean-field equilibrium over time. We have extended the theoretical results in the specific setting of wholesale power markets with RTP and prosumers, hence contrasting our approach to heuristic agent-based simulations.

## Results

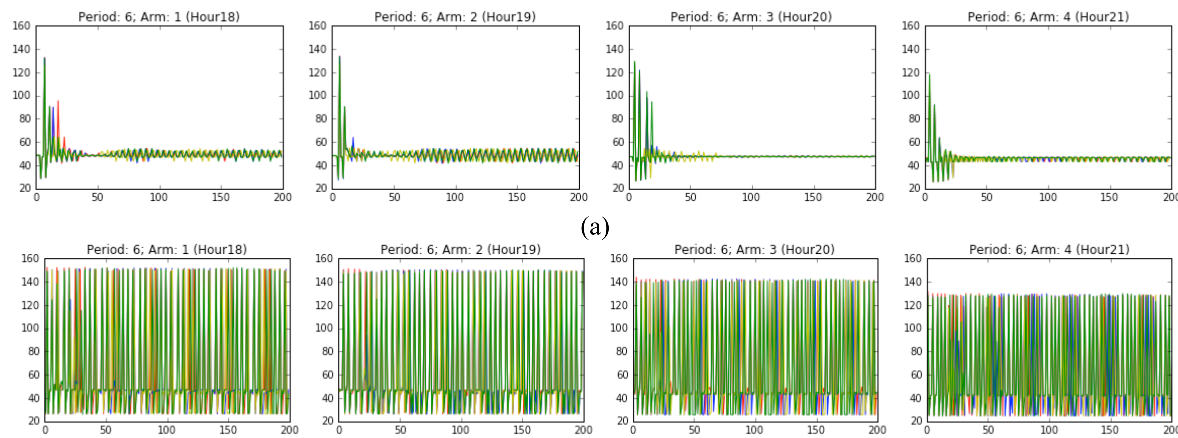
[1] Gummadi, R., Johari, R., Schmit, S., & Yu, J. Y. (2013). Mean field analysis of multi-armed bandit games. Working paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2045842. Last accessed: December 17, 2018.

We use an 8-zone simplified power system representing ISO-NE, as shown in Figure 1(a), together with the 12 transmission lines (each of 1,200 MW capacity). There are 76 fossil fuel-fired generators in the system. Each zone has two types of loads, fixed and flexible. Aggregated (over the 8 zones) fixed load in a day is shown in Figure 1(b). Each zone is assumed to have 600 prosumers with both flexible demand and distributed generation, sampled from a Beta distribution of factors (2,2) and (2,4) (in MW), respectively. We let the simulation run for 200 days, and each day is divided into 6 periods, each consisting of 4 consecutive hours. An agent then tries to pick an hour from the 4-hour period to consume electricity or send electricity to grid. Figure 2 shows the realized RT prices in Period 6 (6 PM – 9 PM) of the Boston zone. It can be seen that while initially oscillating, RT prices under the MAB-game approach quickly converge to a steady state; while huge variations are exhibited under naïve-response. Realized real-time demand show similar patterns. Figure 3 shows the total dispatch cost and congestion cost to further demonstrate the effectiveness of the MAB-game approach over naïve response.



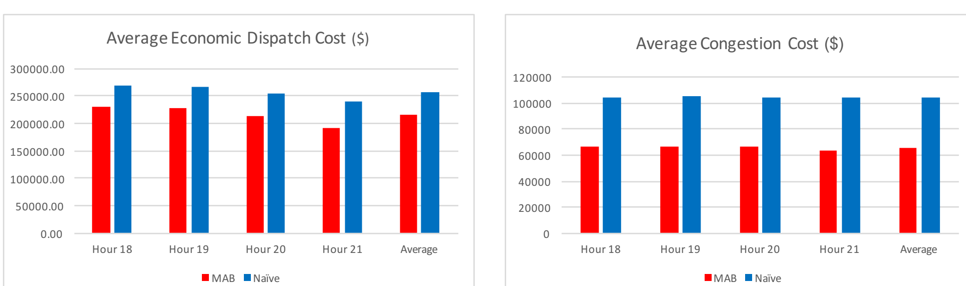

(a) (b)

Figure 1. (a) 8-Zone ISO-NE Test System; (b) Base load profile of 24 hours of the ISO-NE system.




(b)

Figure 2. Hourly real-time (RT) prices of Boston (MW/$) of Period 6: (a) MAB-game; (b) Naïve-response




(a) (b)

Figure 3. Average system costs ($) of Period 6: (a) Economic dispatch costs; (b) Congestion costs.

## Conclusions

In this work, we proposed a decentralized approach to integrate DERs into wholesale power markets under RTP. The weak assumptions on agents' rationality and the scalability of the MAB game make it well-suited for practical implementation. Simulation results demonstrated significant benefits of our approach in reducing price volatility and total dispatch costs when compared to naïve-response approach.