Multi-Armed Bandit Approaches for Real-Time Electricity Pricing with Grid Reliability Constraints

Nathaniel Tucker[1,†], Ahmadreza Moradipari[1], and Mahnoosh Alizadeh[1]
[1]: Department of Electrical and Computer Engineering
University of California, Santa Barbara
Santa Barbara, CA, USA 93106
[†]: Corresponding Author Email: nathaniel_tucker@ucsb.edu

***Index Terms*—Constrained optimization, distribution network, multi-armed bandit, real-time pricing, reinforcement learning**

I. Paper Summary

In order to integrate the increasing volume of stochastic renewable generation in the modern power grid, distribution system operators (DSOs) are exploring various methods to manipulate both residential and commercial loads in real-time. One promising framework known as real-time pricing (RTP) has gained popularity because of its ability to shape electricity demand by exposing customers to time varying prices. However, impediments of RTP are that DSOs do not have access to information on how customers respond to price signals and the customers price responses are stochastic and time varying [1].

In this paper, we consider the aforementioned real-time electricity pricing problem faced by a DSO attempting to manipulate the demand of its customers. Moreover, the DSO wants to *passively* learn (i.e., only utilizing past responses to price signals) the customers' price response models while selecting cost-minimizing daily electricity price signals. Contrary to previous real-time pricing methods that attempt to learn customer price sensitivity models [2]–[6], our methods additionally consider realistic power system constraints, e.g., nodal voltage, transformer capacities, and line flow limits during the run of the learning algorithm. In real distribution systems, it is critical that these constraints are satisfied at every time step to ensure customers receive adequate service and to avoid potential grid failures [7], [8]. When implementing an RTP method, the DSO must ensure that the selected price signals do not lead to constraint violations, even without sufficient knowledge about how customers respond to price signals (i.e., in early learning stages).

To achieve these goals, we make use of the multi-armed bandit (MAB) framework, a well-known problem in reinforcement learning, to select effective price signals while gathering knowledge about the customers' price sensitivities. Specifically, we present two **modified** heuristics akin to Thompson sampling (TS) and upper-confidence bound (UCB) to tackle the polarizing tradeoff between exploration of untested price signals and exploitation of well-performing price signals while ensuring grid reliability. It is important to note that standard bandit heuristics cannot guarantee that the reliability constraints are upheld during the learning procedure, so we present modified versions while retaining the fundamental principles they are based on. Accordingly, we provide a discussion on reliability guarantees for each of the modified learning procedures, a discussion on the regret performance of each heuristic, and extensive simulation results highlighting the strengths of each method.

II. Main Contributions

The main contributions of this paper are as follows:

- We use the multi-armed bandit framework to model the effects of the stochastic and unknown nature of customers' price responses.
- Our problem model takes into account realistic grid reliability constraints that are critical for daily operation.
- We present two modified heuristics based on Thompson sampling and upper-confidence bound as solutions to the reliability constrained learning and pricing problem.
- We provide discussion on the performance of each pricing method, discussion on the reliability guarantees of each heuristic, and a large-scale case study demonstrating the efficacy of each.

References

[1] C. Eid, E. Koliou, M. Valles, J. Reneses, and R. Hakvoort, "Time-based pricing and electricity demand response: Existing barriers and next steps," *Utilities Policy*, vol. 40, pp. 15 – 25, 2016.

[2] A. Moradipari, C. Silva, and M. Alizadeh, "Learning to dynamically price electricity demand based on multi-armed bandits," in *2018 IEEE GlobalSIP*, Nov 2018, pp. 917–921.

[3] L. Jia, Q. Zhao, and L. Tong, "Retail pricing for stochastic demand with unknown parameters: An online machine learning approach," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2013, pp. 1353–1358.

[4] K. Khezeli and E. Bitar, "Risk-sensitive learning and pricing for demand response," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6000–6007, 2017.

[5] P. Li, H. Wang, and B. Zhang, "A distributed online pricing strategy for demand response programs," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 350–360, 2017.

[6] V. Gmez, M. Chertkov, S. Backhaus, and H. J. Kappen, "Learning price-elasticity of smart consumers in power distribution systems," in *2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm)*, Nov 2012, pp. 647–652.

[7] E. DallAnese, K. Baker, and T. Summers, "Chance-constrained ac optimal power flow for distribution systems with renewables," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3427–3438, 2017.

[8] R. Mieth and Y. Dvorkin, "Data-driven distributionally robust optimal power flow for distribution systems," *IEEE Control Systems Letters*, vol. 2, no. 3, pp. 363–368, 2018.