

# Distributed Learning Algorithms for Scheduling Games in the Future Smart Grid

Antimo Barbato<sup>†</sup>, Antonio Capone<sup>†</sup>, Lin Chen<sup>‡</sup>, Fabio Martignon<sup>‡¶</sup> and Stefano Paris<sup>§</sup>

<sup>†</sup> DEIB

<sup>‡</sup> LRI

<sup>§</sup> LIPADE

Politecnico di Milano

Paris-Sud University

Paris-Descartes University

{antimo.barbato,antonio.capone}@polimi.it

{lin.chen,fabio.martignon}@lri.fr

stefano.paris@parisdescartes.fr

¶ Institut Universitaire de France (IUF)

**Abstract**—Game theory is a key analytical tool to design Demand-Side Management (DSM) systems, since it can be used to model the complex interactions among the independent actors of the smart grids. In this paper, we propose two learning algorithms to enable the players of game theoretic DSM frameworks to autonomously converge to the Nash equilibria of the game, and we evaluate their performance based on real instances of the problem.

## I. INTRODUCTION

Demand-side management systems represent an efficient method to control and schedule the consumers' appliances with the aim of improving the efficiency of smart grids. Specifically, these solutions can be applied to shift the users' demand from peak to off-peak periods, therefore reducing the need for generation, transmission and distribution capacity, as well as the power grids investments.

In the field of DSM systems, game theoretic methods have gained increased momentum, since they can be used to model and study the interactions among the independent rational users of power grids. In this case, the demand management problem is formulated as a game, where the players are the consumers or the appliances themselves, the strategies are the schedules of devices and the utility functions are the energy bills [1]. The goal of these methods is to drive the system to equilibria that improve the efficiency of the power grid. However, converging to the game equilibria is a non-trivial challenge and learning algorithms are required to this end [2].

In this paper, we propose two distributed learning algorithms, which enable the consumers to autonomously converge to the equilibria of DSM games through an iterative procedure. Specifically, in these algorithms, the scheduling decision problem of each player is modeled as a Markov chain: each feasible appliance schedule is associated with a state of the chain, and state transition probabilities are updated at every iteration of the procedure, based on the bills of players. At each iteration, all appliances switch to a new schedule with a probability proportional to the cost difference between the actual and cheapest schedules of the previous iteration.

The paper is organized as follows. Section II describes the learning algorithms that we have designed to converge to the Nash equilibrium of the DSM game. Numerical results are provided in Section III and conclusions are drawn in Section IV.

## II. DSM LEARNING ALGORITHMS

In this paper, we consider a generic smart grid model in which each consumer  $h$  of a group of residential users,  $\mathcal{H}$ ,

has a set of appliances,  $\mathcal{A}$ , that have to be scheduled over a 24-hour time period divided into a set,  $\mathcal{T}$ , of time slots. Each device  $a$  of consumer  $h$ , which is characterized by a load profile  $l_{ah}$ , must be executed only once within the time window  $[ST_{ah}, ET_{ah}]$ . The energy tariff is defined based on a real-time pricing scheme. Specifically, since the higher the demand of electricity, the larger the capacity of grid generation and distribution to install, the electricity price at each time  $t \in \mathcal{T}$ ,  $c_t$ , is defined as an increasing function of the total power demand,  $y_t$ , of the group  $\mathcal{H}$  at time  $t$ .

The appliance scheduling problem can be modeled as a non-cooperative game  $G = \{\mathcal{N}, \mathcal{I}, \mathcal{U}\}$  [3]:  $\mathcal{N} = \mathcal{A} \times \mathcal{H}$  is the players set (player  $n = (a, h)$  is the appliance  $a$  of consumer  $h$ ),  $\mathcal{I} \triangleq \{\mathcal{I}_n\}_{n \in \mathcal{N}}$  is the set of strategies which correspond to the appliances schedule and  $\mathcal{U} \triangleq \{U_n\}_{n \in \mathcal{N}}$  is the set of utility functions that coincide with the devices electricity bills. In this DSM game, each player (appliance)  $n$  chooses its strategy  $\mathcal{I}_n$  to minimize its cost  $U_n$ .

In order to enable players to converge to the Nash equilibria of the game, we have designed two learning algorithms based on the proportional imitation rule [4]. In these algorithms, which are defined as iterative procedures, the schedule selection problem of each appliance  $n \in \mathcal{N}$  is modeled as a Markov chain (see Figure 1) in which each state  $s \in \mathcal{S}^n$  is associated with a possible feasible schedule of the appliance  $n$ . At each iteration (i.e., day)  $k$ , the appliance  $n$  selects a certain schedule/state  $s \in \mathcal{S}^n$  which gives to  $n$  a certain utility (i.e., bill)  $U_n^k$ . This bill, together with the potential bills that  $n$  would have paid if it had executed any other schedule, is used to update the states transition probabilities of the chain  $P(s_i, s_j)$  with  $s_i, s_j \in \mathcal{S}^n$ . At the next iteration  $k + 1$ , the process randomly moves to a new state  $s' \in \mathcal{S}^n$  based on the transition probabilities, and the appliance  $n$  receives a new bill  $U_n^{k+1}$ .

In order to define the transition probabilities, we have designed two policies: Two-State (TS) policy and Multi-State (MS) policy. In the first case, in each iteration  $k$ , the initial Markov chain of Figure 1 is reduced to a two-state chain, since only two possible alternatives are considered to select the device schedule for the current day: keeping to use the old schedule or switching to the cheapest schedule among the feasible ones. In order to foster the change whenever the alternative schedule is more convenient, the transition probabilities are updated according to the difference between the corresponding schedules' electricity prices. On the other hand, in the MS case, the algorithm considers all the cheaper schedules among all the feasible alternatives besides the old schedule and the transition probabilities are properly defined so that the lower the ratio between the old and the cheaper schedule, the higher the

probability to change and use such more convenient schedule.

### III. NUMERICAL RESULTS

The proposed learning algorithms have been tested on realistic instances of the DSM problem [5], [6]. Specifically, we have considered a group of 100 consumers, each one having 4 shiftable devices out of 11 realistically-modeled appliances (i.e., shiftable devices: washing machine, dishwasher, boiler, vacuum cleaner; fixed devices: refrigerator, purifier, lights, microwave oven, oven, TV, iron). As for the load scheduling flexibility (i.e., the size of the  $[ST_n, ET_n]$  time-window), we have defined three scenarios: no flexibility, in which the appliances schedule is fixed, low flexibility and high flexibility in which, respectively, 3 and 8 different possible schedules are randomly set for each device.

Numerical results are presented in Figure 2, in which we compare the aggregated users' bill obtained with the proposed learning algorithms with the one associated with the Nash equilibrium of the DSM game, in order to verify the convergence of the learning methods. Moreover, we also show the aggregated bill paid by the consumers when the usage of the electric devices is not modified by the DSM system (i.e., no scheduling flexibility).

As one can note from Figure 2, both the TS and MS learning policies converge rapidly to the Nash equilibrium. Specifically, the two-state policy converges slightly more quickly than the multi-state one, which has more alternative schedules to select at each iteration and, therefore, a more extensive exploration phase. Other than the policy, the convergence time depends also on the flexibility of the DSM system: the algorithms converge more rapidly to the Nash equilibrium with a short flexibility level, since, in this case, each appliance has a lower number of strategies to try, therefore requiring a shorter exploration process. However, the shorter convergence time comes at the cost of a worse performance of the DSM system in terms of aggregated bill even if, even with low flexibility, the consumers always achieve lower bills than those obtained with fixed schedules.

### IV. CONCLUSIONS

In this paper, we presented two distributed learning algorithms to enable the consumers to autonomously converge to the equilibria of loads scheduling games. In these methods,

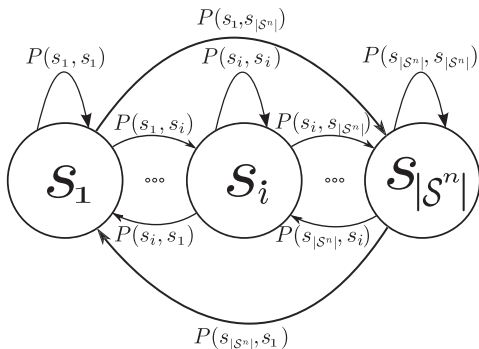
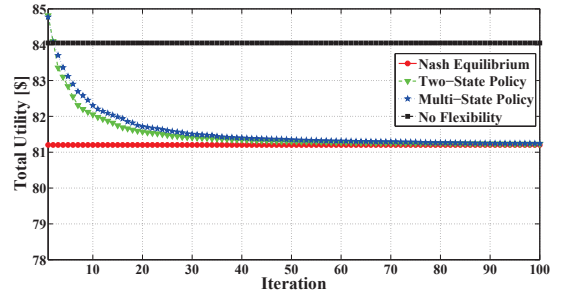
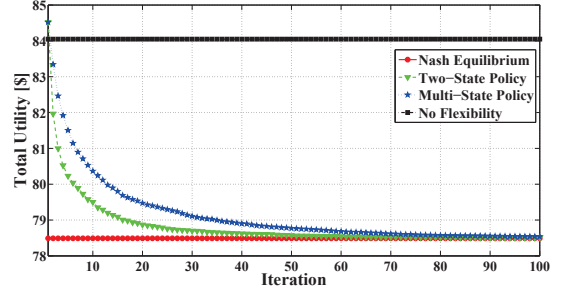


Fig. 1. Markov decision chain of the appliance  $n$  used in the scheduling learning process.



(a) Low Flexibility



(b) High Flexibility

Fig. 2. Aggregated utility with 100 users and different scheduling flexibility levels.

which are defined as iterative processes, each appliance decides autonomously its best schedule. The scheduling decision problem, which is based on the proportional imitation rule, is modeled as a Markov chain, where each feasible appliance schedule is associated with a state of the chain, and the states transition probabilities are updated according to the difference between the paid electricity price and the bills associated with the other schedules.

The performance of the proposed learning methods have been evaluated based on realistic instances of the DSM game. Numerical results have shown that these algorithms quickly converge to stable Nash equilibria, which lead to cheaper bills than those obtained without the proposed DSM framework.

### REFERENCES

- [1] A. Mohsenian-Rad, V. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid," *IEEE Trans. on Smart Grid*, vol. 1, no. 3, pp. 320–331, 2010.
- [2] D. Fudenberg, *The theory of learning in games*. MIT press, 1998, vol. 2.
- [3] A. Barbato, A. Capone, L. Chen, F. Martignon, and S. Paris, "A distributed demand-side management framework for the smart grid," 2014, preprint, <http://arxiv.org/abs/1405.1964>.
- [4] S. Iellamo, L. Chen, and M. Coupechoux, "Proportional and double imitation rules for spectrum access in cognitive radio networks," *Computer Networks*, vol. 57, no. 8, pp. 1863–1879, 2013.
- [5] ECORET Project, Official web site (ITA), [http://www.rse-web.it/progetti.page?RSE\\_originalURI=/progetti/progetto/documento/178/312827&objId=178&typeDesc=Rapporto&RSE\\_manipulatePath=yes&docIdType=1&country=ita](http://www.rse-web.it/progetti.page?RSE_originalURI=/progetti/progetto/documento/178/312827&objId=178&typeDesc=Rapporto&RSE_manipulatePath=yes&docIdType=1&country=ita), Sept 2014.
- [6] MICENE Project, Official web site (ITA), [http://www.eerg.it/index.php?p=Progetti\\_-\\_MICENE](http://www.eerg.it/index.php?p=Progetti_-_MICENE), Sept 2014.