Energy Syst (2012) 3:61–76
DOI 10.1007/s12667-012-0054-0

ORIGINAL PAPER

An adaptive multi-agent-based approach to smart grids control and optimization

Marco Carvalho · Carlos Perez · Adrian Granados

Received: 5 December 2011 / Accepted: 28 January 2012 / Published online: 14 February 2012

Abstract In this paper, we describe a reinforcement learning-based approach to power management in smart grids. The scenarios we consider are smart grid settings where renewable power sources (e.g. Photovoltaic panels) have unpredictable variations in power output due, for example, to weather or cloud transient effects. Our approach builds on a multi-agent system (MAS)-based infrastructure for the monitoring and coordination of smart grid environments with renewable power sources and configurable energy storage devices (battery banks). Software agents are responsible for tracking and reporting power flow variations at different points in the grid, and to optimally coordinate the engagement of battery banks (i.e. charge/idle/discharge modes) to maintain energy requirements to end-users. Agents are able to share information and coordinate control actions through a parallel communications infrastructure, and are also capable of learning, from experience, how to improve their response strategies for different operational conditions. In this paper we describe our approach and address some of the challenges associated with the communications infrastructure for distributed coordination. We also present some preliminary results of our first simulations using the GridLAB-D simulation environment, created by the US Department of Energy (DoE) at Pacific Northwest National Laboratory (PNNL).

M. Carvalho (✉)
Florida Institute of Technology, 150 W. University Blvd., Melbourne, FL 32901, USA
e-mail: mcarvalho@fit.edu

C. Perez · A. Granados
Institute for Human and Machine Cognition, 15 SE Osceola Ave, Ocala, FL 34471, USA

C. Perez
e-mail: cperez@ihmc.us

A. Granados
e-mail: agranados@ihmc.us

Keywords Multi-agent systems · Smart grids · Reinforcement learning · Renewable energy management · Energy storage devices

1 Introduction

One of the key benefits enabled by smart grids is the ability to manage power consumption and distribution more efficiently, and to support two-way information and power flows in the grid. By integrating information technology with advanced monitoring and control at all levels in the system, future infrastructures will be smarter, more effective, and likely more resilient to localized failures or malicious attacks. This capability also enables the potential for distributed power management, or self-managing infrastructures that can balance cost, service and resilience based on current demand, policies, and environmental conditions.

The increasing acceptance of smart grid and renewable energy technologies brings new capabilities and opportunities for future power systems. However, as it is often the case, new system capabilities also bring new challenges for the design and management of power generation and distribution infrastructures. Renewable energy sources enable a number of advantages that include, for example, long term cost reductions and the potential reduction of carbon emissions. However, they are generally dependent on environmental conditions and other external factors that may affect their reliability, and possibly the overall dependability of the power system.

The challenge then becomes to design new systems and infrastructures that can leverage the potential benefits provided by distributed and renewable energy sources, while maintaining acceptable levels of reliability in the overall system, specially in peak hours and critical operation conditions.

One way to mitigate the problem is to compensate for the renewable source fluctuations by controlling the real-time energy demand through automated demand response. Demand response mechanisms usually imply providing a price signal, or incentive to smart grid devices in order to allow them to properly schedule non-critical loads and optimize their costs [4, 12, 21]. Dam et al. [9], for example, proposed an expert system-based approach to control local energy loads based on pricing information and local load management policies.

O'Neill et al. [18] proposed CAES (Consumer Automated Energy Management System), a reinforcement learning-based approach to demand control, primarily based on the estimation of the impact of future energy prices and consumer decisions on residential energy management. Demand response mechanisms are very effective for accommodating relatively short-term changes in energy price and availability, but they must be carefully balanced with customer preferences and requirements to maintain a perceived quality of service. That may, in some cases, limit their application to mitigate fast, short term transients in energy supply.

An alternative approach to address the problem is to provide a system-level stabilization mechanism that can accommodate renewable supply transients without minimum impact to the perceived quality of service to the end-user. These strategies would consist of stabilizing the supply of energy by using distributed storage mechanisms to mitigate the short term supply transients. By properly choosing when to charge and discharge distributed storage devices, such as batteries, the system can compensate for both, fluctuations in demand and in supply (power generation). With short term fluctuations mitigated, longer-term variations in the system could still be managed through fall-back demand response mechanism, if necessary.

This approach, however, also faces its own set of challenges. First, there is still the need for demand response mechanisms, because the charging of the storage devices must be scheduled in a way that leverages the spare renewable energy, minimizing any use of non-renewable energy sources. Second, power losses in the lines increase with distance, therefore, it is desirable to discharge the storage devices that are closer to the areas of high demand. And third, using the distributed storage to handle unpredictable disruptions might cause an opportunity cost if the storage devices are not properly charged for other programmed disruptions.

In order to optimally manage the storage devices in coordination with the renewable power sources, a number of optimization and optimal control strategies have been proposed [6]. For example, in [17] the authors proposed an optimal bidding strategy to combine renewable power generation with energy storage. The proposed approach is centralized, and the problem is formulated as a continuous Markov decision process. In their approach, the power storage capabilities are modeled as a risk mitigation component for the uncertainties in price and supply, of renewable energy sources. In [24], a rule-based approach is proposed for battery-based energy storage and dispatching. The goal is to maintain energy supply levels, for intermittent data source environments. The approach is also centralized, and based on a fixed rule-based controller to smooth the transients of renewable energy sources based on forecasted solar and wind conditions. Other related approaches have also been proposed and discussed in the literature, including [1, 2, 13], and [8].

More recently, multi agent-based coordination strategies have been introduced to address the problem [20]. For example, in [28] the authors propose a multi-agent game theoretic approach to minimize the probability of peaks in energy demand, when multiple micro-storage devices are trying to charge from energy produced by the grid. In [19], an agent-based homeostatic control strategy is proposed for grid stability using storage devices and learning software agents for smart grid management. In [16], a multi-agent approach is proposed for the distributed coordination of energy production and storage in a hybrid system connected to the grid. In that work, agents directly affected individual convertors connected to the grid by setting, for example, their control mode and current reference. Wei et al. [29], built on the work of [25, 27] and [10] to propose an agent-based algorithm for energy storage management where individual agents shared used short term (day-ahead) energy price predictions for store allocation. For additional information, a more theoretical discussion of the benefits and challenges associated with agent-based optimization and control was discussed in [27]. In this paper, we propose a supply control mechanism using reinforcement learning to manage the charging and discharging of storage devices (i.e. battery banks), as a function of the system load and available sources. The goal is to minimize the costs associated with drawing additional energy from the main grid, due to temporary variations in the energy generated by the renewable sources [26].

2 An illustrative scenario

The scenario shown in Fig. 1 illustrates a simplified version of a smart grid configuration with the same set of components we consider in this work. The scenario consists of a small feeder with one connection to the main grid (top input node), distributed energy loads (houses), distributed energy storage (batteries) and a renewable energy source. There is a meter located at the top of the feeder to keep track of the overall power exchanged with the main grid. The solar array at the top of the feeder provides a renewable source of energy to the system and is also connected to a separate meter that tracks power being generated by the solar array.

The batteries are distributed across the feeder and are also connected to individual meters to keep track of the power produced or consumed by each battery, as well as the load on the branch where the battery is located. While the proposed scenario includes only three batteries, placed in different branches of the feeder, the algorithm does not constrain the number or location of the batteries, as long as the effects of their states are perceived in the set of meters being monitored.

One important difference between the proposed scenario and some of the previous settings considered for multi-agent energy management is the distribution of storage devices. Conventionally in the literature, energy storage devices are considered at the end-user side, that is, in individual homes and buildings. These are relatively low capacity devices used to temporarily store spare renewable energy, for example from solar, or wind sources. At the local level, the optimization of energy storage is also

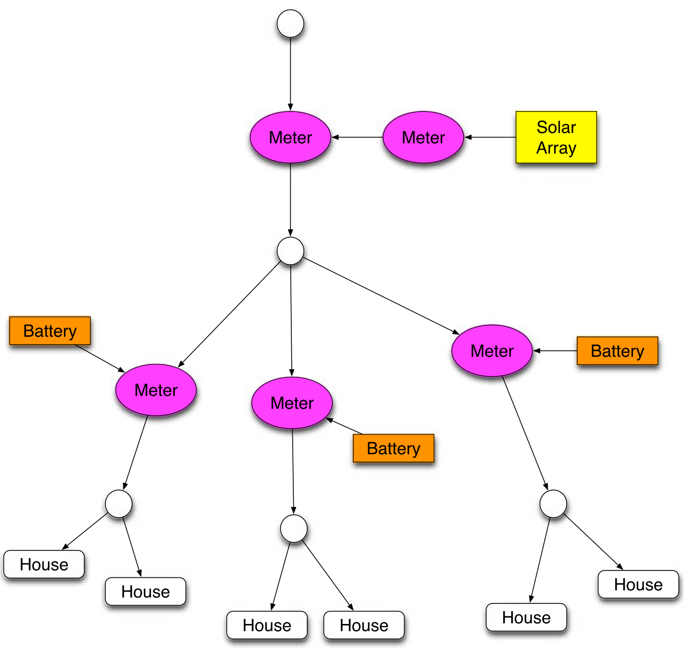

Fig. 1 A simplified illustrative scenario

important, and may take into account varying costs of power (from the main grid), and power utilization profiles. In our scenario, we envision larger scale battery banks deployed in specific (and fixed) locations in the micro-grid. While the optimization of battery placement is in itself an interesting research problem, it is not part of the scenario we are addressing. In our scenario, battery banks are pre-located in the grid, with only a limited control interface to set their operational mode as charging, idle or discharging.

The concept of large scale energy storage systems is not new. Early Battery Energy Storage Systems (BESS) were proposed in the late 70's as part of a US renewables energy initiative. Numerous related projects follow, including the Exploratory Battery Development Testing Program (ETD) in the early 80's, the Utility Battery Storage Program (USB) in the early 90's, and more recently the Department of Energy's Energy Storage System (ESS) program, which includes not only batteries, but also other alternative energy storage technologies such as flywheels and compressed air energy storage (CAES). The potential for large scale storage devices, for example for window power, were discussed in [15]. In [11], a review of the state of the art battery energy storage is provided, as well as their economic viability and impact on the power system operation. A comparative analysis of different large scale energy storage technologies is presented in [23].

While different technologies have very different characteristics and capabilities, for the purposes of our work we consider a simplified battery (or battery bank), as a storage device. Utilization costs and charge/discharge rates are fixed, and all battery banks in the scenario are considered identical. The scenario used for the experimental results consists of a larger feeder than the one illustrated in Fig. 1. The feeder contains 15 houses and 3 batteries, and it was based on the IEEE 13-Node radial distribution test feeder [14].

3 Learning an optimal strategy for dynamic battery allocation

In the proposed scenario, the feeder connects with the main grid in one single point, and is able to draw the necessary energy to accommodate the dynamic load of the system. However, the costs of energy drawn from the main grid are significantly higher when compared with energy costs obtained from renewable energy sources. For the purpose of reducing overall energy costs, a solar array was installed and connected to the main meter. In addition to the solar array, three sets of batteries were also deployed in the system to mitigate the variations of energy output from the solar array due to cloud transients or other external effects.

In this work, a dynamic programming approach is proposed for learning, from experience, an optimal policy for engaging each battery based on measurements from the meters. The goal is to minimize the costs associated with energy utilization from the main grid. Each battery can be independently set in one of three states: charging, idle and discharging, and the goal of the learning algorithm is to determine the state of each battery given the readings from the different meters. Software agents are associated with each battery, and communicate with each other to share information and coordination actions. There are many options for the communications infrastructure [7], and specific implementations are generally driven by economic, geographic, and existing infrastructure constraints.

A reinforcement learning system is characterized by a policy, a reward function and a value function [22]. Policies are the core of the reinforcement learning task and are often stochastic. They define the way the agent behaves at a given time and are basically mappings from states of the environment to the actions that ought to be taken when the system is in those states.

A reward or cost function defines the goal in a reinforcement learning task. The main goal of the reinforcement learning system is to maximize the total reward, or minimize the total cost in the long run. The value function specifies, for every perceived state of the environment, which action is good in the long run. For each state-action combination, the value function indicates the desirability of taking the action when the system is in the given state. The value function is usually identified by $Q(s, a)$, where s is the state, and a is the action.

Reinforcement learning algorithms rely, in general, on the estimation of a value function. The method that we use here for learning the value function is know as Q-learning. After taking an action a while being in state s and observing the resulting reward r and the new state s, a Q-learning agent will update the value function using the following formula:

$$Q(s, a) \leftarrow Q(s, a) + \alpha\Big[r + \gamma \max_{a'} Q(s', a') - Q(s, a)\Big], \tag{1}$$

where α is known as the learning rate; and γ is known as the discount rate and determines the present value of future rewards.

The reinforcement learning agent will construct a policy using the Q function. The policy is stochastic, giving different probabilities to each action according to the estimated value of the action in the current state.

Some of the most common strategies for policy discovery include ε-greedy and *softmax*. An ε-greedy strategy selects the action with the highest value with probability $(1 - \varepsilon)$, and splits the remaining probability uniformly among the other actions. A *softmax* strategy varies the probabilities as a graded function of the estimated value. The most common softmax method uses the Boltzmann distribution, and chooses action a with probability:

$$P_s(a) = \frac{e^{Q(s,a)/\tau}}{\sum_{b \in K} e^{Q(s,b)/\tau}}, \tag{2}$$

where K is the set of possible actions and τ is a positive parameter called the *temperature*.

High temperatures cause the actions the be all nearly equiprobable. Low temperatures cause greater difference in selection probability for actions that differ in their value estimates. The problem with this strategy is that all actions need to be explored at least once to be able to compute the probability. So, a strategy known as *Exp3* is used to help bootstrap the function. Basically, what it does is to assign a uniform distribution over the unexplored actions, and then combines it with the Boltzmann distribution.

The state of the system can be modeled with different degrees of granularity. A possible approach consists on using as state the measurements on the top level meter and the solar array meter. This approach requires factoring out the power generated or consumed by the batteries from the top level meter. In this case, the actions do not change the state of the system.

An alternative approach consists on using the batteries meters and battery states (charging, discharging, idling) in addition to the top level meter and solar array meter. By using this approach, there is no need to factor out the power generated or consumed by the batteries because the state of the batteries is already included as part of the state.

The measurements in the meters will be divided into five levels: high negative, low negative, neutral, low positive and high positive. The negative values occur when the solar array is producing more energy than the system can consume. The positive values occur when the energy produced by the solar array is not enough to feed the system. A neutral level occurs when the power measurement is close to zero, meaning that the system is self-sufficient.

An action of the system is the combination of three possible commands for each battery: charge, discharge or idle. The charge and discharge actions will be fixed at a given power level, but nothing forbids from using several levels of charge/discharge power.

The cost function has three components:

- $|P|$: the absolute value of the power consumed by the system at the top level meter.
- The cost of consuming power from the grid (when $P > 0$).
- The cost of charging/discharging the batteries.

The cost function is defined by:

$$Cost = \alpha \cdot |P| + \beta \cdot H(P) \cdot P + \gamma \cdot B_{cost}, \tag{3}$$

where P is the power level; α, β and γ are weight multipliers for the cost components; B_{cost} is the cost associated with the batteries, and $H(P)$ is the function defined by:

$$H(P) = \begin{cases} 0, & P < 0 \\ 1, & P \geq 0 \end{cases} \tag{4}$$

3.1 Estimating the battery costs

The cost of charging/discharging the batteries is based on the expected life of the battery. In our formulation, we disregard the investment costs in the storage device and their amortization. We focus instead, solely in the costs associated with the use of the device, effectively based on the expected number of cycles for individual batteries. The overall battery cost is the sum of the costs for all batteries, and the usage cost for individual batteries is given by the following:

$$B_{cost} = |P_B| \cdot \alpha \cdot (1 - C_B) + \big(1 - H(C_s)\big) \cdot |P_B| \cdot \beta, \tag{5}$$

where α and β are multipliers for the cost components, P_B is the power step for the battery when charging/discharging, C_B is the battery charge (0%–100%), and C_s is the following function:

$$C_s = \begin{cases} 1, & \text{if charging} \\ 0, & \text{if idling} \\ -1, & \text{if discharging} \end{cases} \tag{6}$$

3.2 Policy learning

A well-known policy learning algorithm is the Exp3, which stands for (Exponential-weight algorithm for Exploration and Exploitation) [3]. In our formulation we use a variation of the Exp3 action selection strategy will be used. It combines the Boltzmann distribution over the value of actions already explored, with a uniform distribution over the unexplored actions. Using this strategy, the probability of selecting action a when the system is in state s is defined by:

$$P_s(a) = (1-\alpha) \cdot \frac{e^{Q(s,a)/\tau}}{\sum_{b \in K} e^{Q(s,b)/\tau}} + \frac{\alpha}{|K|}, \tag{7}$$

where K is the set of possible actions, and α is the exploration probability.

4 Experimental results

To evaluate the proposed approach a simple grid scenario, illustrate in Fig. 2, was implemented and simulated in GridLAB-D [5]. The reinforcement learning agent was implemented as a separate module to GridLAB-D, loaded with the scenario at runtime. The state was modeled using the state of all the meters and the state of the batteries. This model was chosen to avoid having to adjust the power in the top meter to factor out the power produced, or consumed, by the batteries.

The scenario runs 2 hours of simulated time, during which the solar flux given to the solar panels changes every 5 minutes between 0, 500 and 1000. The reinforcement learning agent checks the state of the system every 100 ms to decide if an action must be taken or not. We assume that some of the meters are monitored by a local agent that responds to state information requests and control commands sent from a centralized control point. In our example, the centralized control point is an agent located at the entry-point of the feeder. However, that role could also be played by a remote component to the system, or the utility company itself. The effects of communication delays between individual agents and the main coordination node were also simulated. Figure 3 illustrates the simulated communications topology.

First we will show the results of running the scenario under an idealistic communications network with a maximum delay below 10 ms. Under these conditions, all measurements performed by the meters and the batteries that are sent to the agent take less than 10 ms to reach their destination. In the same manner, all actions taken

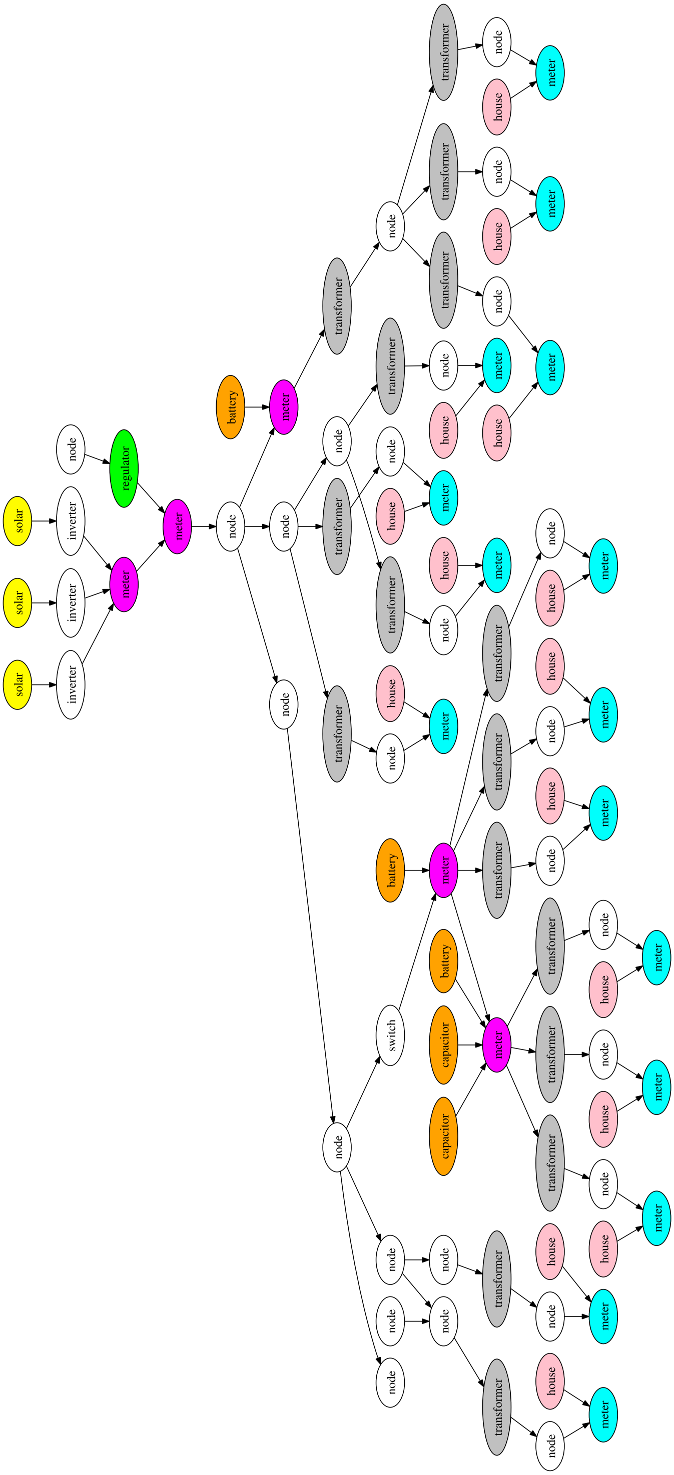

Fig. 2 Simulated Feeder scenario

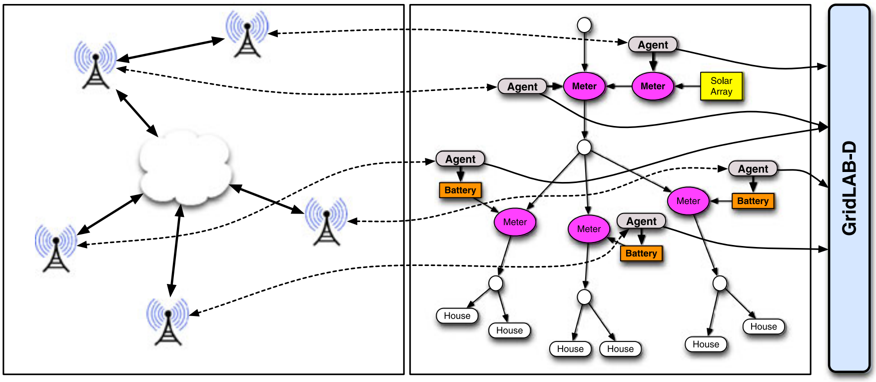

Fig. 3 Simulating the effects of communication delays on distributed meter monitoring

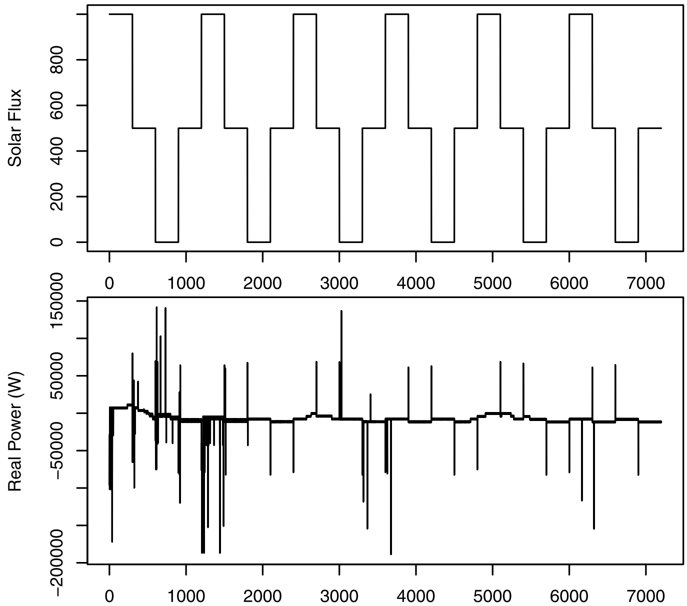

Fig. 4 Results under an idealistic communications delay below 10 ms

by the agent take less than 10 ms to be communicated to the batteries. Then, we will show how the communications delay can affect the system.

Figure 4 shows how the system succeeds in keeping the real power near zero under an idealistic communications infrastructure with communication delays under 10 ms. At the beginning of the scenario, the system struggles to keep the power close to zero because it is exploring possible solutions for each state change, that is, the system is learning an optimal policy for the different operational conditions. Once learned, similar conditions can be properly addressed, as the system has already developed an optimal strategy for the different conditions. As illustrated in Fig. 4 the exploration for new strategies happens when there is a change in the solar flux, under a new system state.

The results illustrated in Fig. 4 show how the system can quickly adapt to variations in power generated by the solar arrays. However, the results are based on idealistic communication conditions. The default delay of 10 ms is 10 times smaller than the rate at which the agent performs an action (100 ms). Under these assumptions, the agent will almost always have an updated view of the state of the system, and it can expect that the effects of its actions will be visible in the following iteration. However, a more realistic representation of the communication delays could not only include much larger latencies in the state propagation, but also very different delay profiles for different meters.

We then explore the effects of two types of delays in the state and control propagation. In the first case, we consider a general or systemic delay that equally affects the perception of the learning agent regarding the state of all sensors. This type of delay might be representative of deployments in which the control is performed at a location relatively far (from a communications perspective) from all the entities being monitored and controlled. Under such conditions, the delay tends to similarly affect all meters. In the second case, we consider the effects of a more realistic model, in which each meter and battery has different communication delays (i.e. variable delay).

The effects of the communications infrastructure on the system are illustrated in Fig. 5, which shows the results for the systemic delay on all links. The figure also shows the plot for the solar flux and the scenario with a 10 ms delay, just to clarity and to facilitate the comparison. Figure 5 also shows the plots for the average delays of 30, 50, 100, 300 and 500 ms.

The simulated delays are random, and normally distributed with a standard deviation of 10% the average delay. The learning agent update rate is still 100 ms, so the effects of the actions of the agent under the 300 and 500 ms delay will only be noticed 6 or 10 iterations later. The delays are simulated as one way delays (as opposed to round-trip delays), so under a 300 ms, it takes 300 ms for the command to reach the batteries, and another 300 ms to propagate the response to the command. As illustrated in the figure, the system seems to be robust against this type of delay, at least up to 300 ms (3 times the update rate).

The robust behavior of the system can be explained by the fact that even though the learning agent receives the state updates in a delayed fashion, it still keeps a consistent state for the system, as the communications delay is generalized. In this scenario, the effects from an action manifest themselves within the same window for all meters, allowing the system to correlate its actions with the corresponding effects. For the more complicated case, we re-implemented our simulations under a different communications environment, where delays are not the same for all links. Figure 6 shows the results for the variable delay.

The top level meters have a communications delay between 10 and 30 ms, and the batteries and their respective meters have communication delays between 70 and 100 ms. The system, as illustrated in the figures, seems unable to handle this type of delay using an update rate of 100 ms. As the effects of some actions arrive sooner

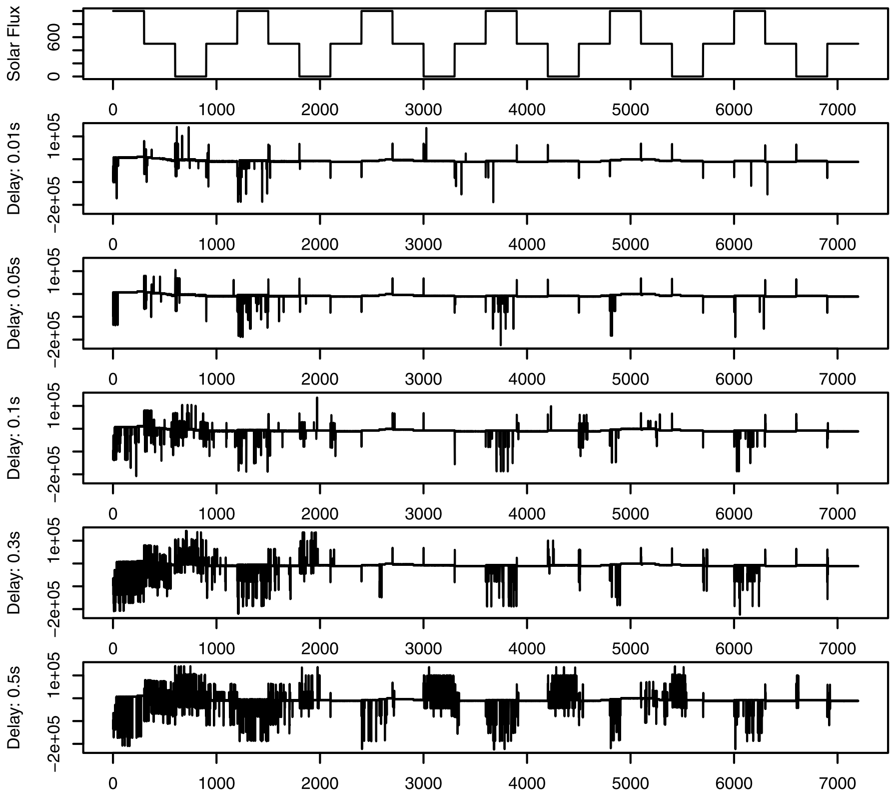

Fig. 5 Effect of a systemic communications delay (10, 30, 50, 100, 300 and 500 ms) with an agent update rate of 100 ms

than others, the system is unable to get a clear picture of the general effect of its commands.

It is possible, however to compensate for the communication delays, as long as they can be estimated. When it is possible to monitor or estimate the delays from all the elements, the system can define an update rate that is higher than the greatest delay. Figure 7 shows the results for different update rates for the same scenario with variable delays. An increase in the update rate from 100 to 120 ms shows little improvement, but an update rate of 150 ms offers better results. An update rate of 200 ms shows results close to the results obtained under idealistic communication delays.

5 Conclusion

The potential for self-management, not only in response to localized failures and attacks, but also as a mean to optimize the response and control strategies for different operational conditions, is of key importance for overcoming the challenges of distributed management and generation in future power systems. In this context,

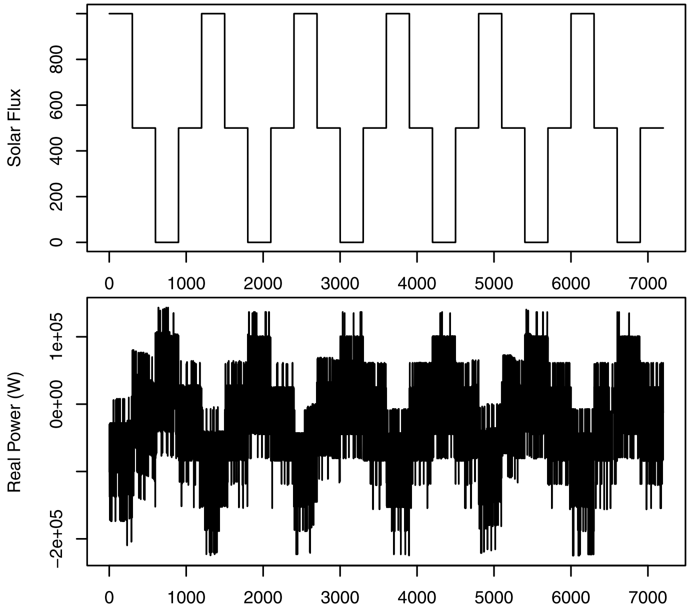

Fig. 6 Effect of a variable communications delay (top level meters have a communications delay between 10 and 30 ms, and the batteries and their respective meters have communication delays between 70 and 100 ms). The agent update rate is 100 ms

reinforcement learning has been shown to be a viable solution for balancing the distributed generation of electricity from renewable and non-renewable sources.

Our adaptive multi-agent approach builds upon a reinforcement learning strategy to control the supply of energy while reducing the cost due to power losses and energy generation from non-renewable sources. It efficiently manages the power and demand response mechanism of the storage devices in the smart grid, and preliminary results in simulation show that the agents are capable of maintaining the demand requirements of the power grid under fixed and variable communication delays.

The implicit assumption of the proposed formulation is that renewable power generated within the feeder (solar panels) is always cheaper than power drawn from the main grid. From that perspective the optimization algorithm focuses on minimizing the energy flow from the main grid to the local feeder, regardless of cost information. Note that minimization in this case, is indeed a maximization of a negative flow—reflecting renewable energy sold back to the main grid.

To take energy costs into account, the proposed approach would have to include another variable representing the relative cost of energy between the main grid renewable sources, which would directly affect the agent's reward. In the simplest case, real-time cost information could be used, which would enable only a reactive optimization of the system. That could be almost directly supported by the current formulation, requiring only a change in the reward function. However, if future prices were taken into account, agents would have to project rewards and lock into a short term strategy based on projected prices. In that case, the state information used for policy

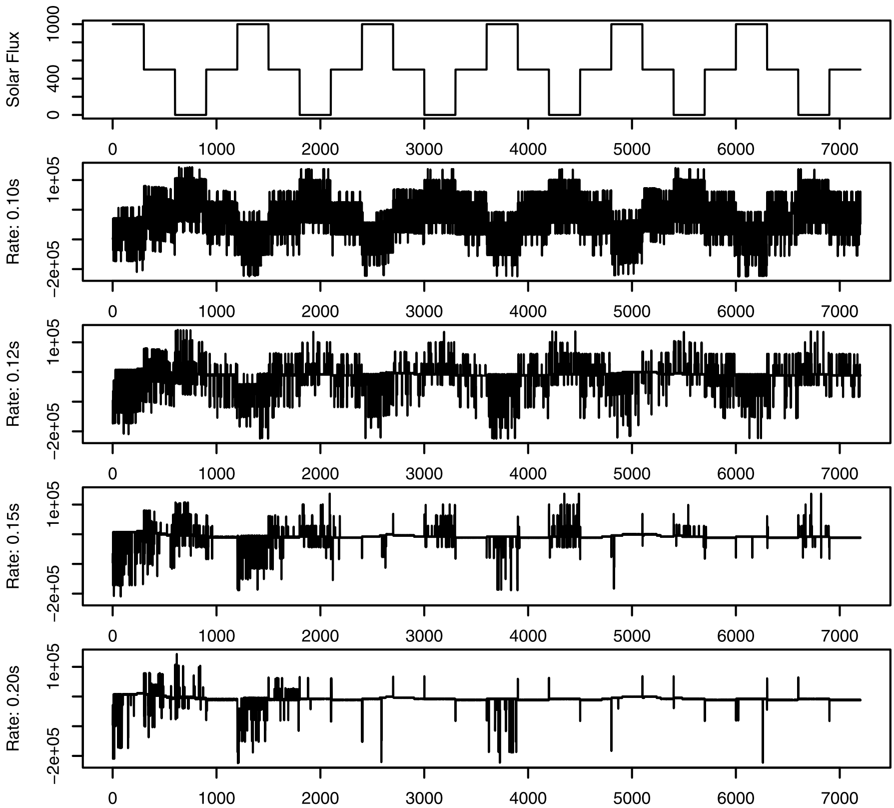

Fig. 7 Result of compensating for the variable communications delay using different update rates

selection (and learning) would have to account not only for current energy costs, but also for projected future costs, as well battery load and discharge times.

As illustrated in our results, the proposed multi-agent system is able to collect and share state information in order to learn optimal strategies for battery allocation under different operational conditions. The characteristics of the communication infrastructure used by the agents significantly influences the effectiveness of the optimization algorithm. However, our results have also shown that such effects can be mitigated by adjusting the update rate of the system in order to improve the overall state syntonization.

References

1. Abbey, C., Robinson, J., Joos, G.: Integrating renewable energy sources and storage into isolated diesel generator supplied electric power systems. In: 13th Power Electronics and Motion Control Conference, 2008. EPE-PEMC 2008, pp. 2178–2183 (2008)
2. Agarwal, Y., Weng, T., Gupta, R.: Micro-systems driving smart energy metering in smart grids. In: DAC'07 (2007)

3. Auer, P., Cesa-Bianchi, N., Freund, Y., Schapire, R.E.: The nonstochastic multiarmed bandit problem. SIAM J. Comput. **32**, 48–77 (2003)
4. Caron, S., Kesidis, G.: Incentive-based energy consumption scheduling algorithms for the smart grid. In: First IEEE International Conference on Smart Grid Communications, (SmartGridComm), 2010, pp. 391–396. IEEE, New York (2010)
5. Chassin, D., Schneider, K., Gerkensmeyer, C.: Gridlab-d: an open-source power systems modeling and simulation environment. In: Transmission and Distribution Conference and Exposition, 2008. D. IEEE/PES, pp. 1–5. IEEE, New York (2008). doi:10.1109/TDC.2008.4517260
6. Chinchuluun, A., Pardalos, P., Enkhbat, R.: Optimization and Optimal Control: Theory and Applications, vol. 39. Springer, Berlin (2010)
7. Cupp, J., Beehler, M.: Implementing Smart Grid Communications (2008). Burns & McDonnell Marketing Communications
8. Dagdougui, H., Minciardi, R., Ouammi, A., Sacile, R.: Optimal control of a regional power microgrid network driven by wind and solar energy. In: 2011 IEEE International Systems Conference (SysCon), pp. 86–90 (2011)
9. Dam, Q., Mohagheghi, S., Stoupis, J.: Intelligent demand response scheme for customer side load management. In: Energy 2030 Conference, 2008. ENERGY 2008, pp. 1–7. IEEE, New York (2008)
10. Dave Cliff, J.B.: Minimal Intelligence Agents for Bargaining Behaviors in Market-Based Environments (1997)
11. Divya, K., Ãstergaard, J.: Battery energy storage technology for power systems: an overview. Electr. Power Syst. Res. **79**(4), 511–520 (2009)
12. Hatami, S., Pedram, M.: Minimizing the electricity bill of cooperative users under a quasi-dynamic pricing model. In: 2010 First IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 421–426. IEEE, New York (2010)
13. Kaldellis, J., Zafirakis, D.: Optimum energy storage techniques for the improvement of renewable energy sources-based electricity generation economic efficiency. Energy **32**(12), 2295–2305 (2007)
14. Kersting, W.: Radial distribution test feeders. In: Power Engineering Society Winter Meeting, 2001. IEEE, vol. 2, pp. 908–912. IEEE, New York (2001)
15. Korpaas, M., Holen, A.T., Hildrum, R.: Operation and sizing of energy storage for wind power plants in a market system. Int. J. Electr. Power Energy Syst. **25**(8), 599–606 (2003)
16. Lagorse, J., Paire, D., Miraoui, A.: A multi-agent system for energy management of distributed power sources. Renew. Energy **35**(1), 174–182 (2010)
17. Löhndorf, N., Minner, S.: Optimal day-ahead trading and storage of renewable energies—an approximate dynamic programming approach. Energy Syst. **1**, 61–77 (2010)
18. O'Neill, D., Levorato, M., Goldsmith, A., Mitra, U.: Residential demand response using reinforcement learning. In: 2010 First IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 409–414. IEEE, New York (2010)
19. Ramchurn, S., Vytelingum, P., Rogers, A., Jennings, N.: Agent-based homeostatic control for green energy in the smart grid. ACM Trans. Intel. Syst. Technol. **2**(4) (2011)
20. Roche, R., Blunier, B., Miraoui, A., Hilaire, V., Koukam, A.: Multi-agent systems for grid energy management: a short review. In: IECON 2010—36th Annual Conference on IEEE Industrial Electronics Society, pp. 3341–3346 (2010)
21. Samadi, P., Mohsenian-Rad, A., Schober, R., Wong, V., Jatskevich, J.: Optimal real-time pricing algorithm based on utility maximization for smart grid. In: 2010 First IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 415–420. IEEE, New York (2010)
22. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge (1998)
23. Suvire, G., Mercado, P., Ontiveros, L.: Comparative analysis of energy storage technologies to compensate wind power short-term fluctuations. In: Transmission and Distribution Conference and Exposition: Latin America (T D-LA), 2010 IEEE/PES, pp. 522–528 (2010)
24. Teleke, S., Baran, M., Bhattacharya, S., Huang, A.: Rule-based control of battery energy storage for dispatching intermittent renewable sources. IEEE Trans. Sustain. Energy **1**(3), 117–124 (2010)
25. Vandael, S., De Craemer, K., Boucké, N., Holvoet, T., Deconinck, G.: Decentralized coordination of plug-in hybrid vehicles for imbalance reduction in a Smart Grid. In: Tumer, K., Yolum, P., Sonenberg, E., Stone, P. (eds.) Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011), pp. 803–810. ifaamas (2011)
26. Venayagamoorthy, G.: Potentials and promises of computational intelligence for smart grids. In: Power & Energy Society General Meeting, 2009. PES'09, IEEE, pp. 1–6. IEEE, New York (2009)

27. Vytelingum, P., Voice, T., Ramchurn, S., Rogers, A., Jennings, N.: Theoretical and practical foundations of large-scale agent-based micro-storage in the smart grid. J. Art. Intel. Res. **42**, 765–813 (2011)
28. Vytelingum, P., Voice, T.D., Ramchurn, S.D., Rogers, A., Jennings, N.R.: Agent-based micro-storage management for the smart grid. In: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: vol. 1 AAMAS '10, pp. 39–46. International Foundation for Autonomous Agents and Multiagent Systems, Richland (2010)
29. Wei, C., Hu, H., Chen, Q., Yang, G.: Learning agents for storage devices management in the smart grid. In: 2010 International Conference on Computational Intelligence and Software Engineering (CiSE), pp. 1–4. IEEE, New York (2010)