

# Optimal Prosumer Decision-Making Using Factored MDPs

**Angelos Angelidakis**

Electronic and Computer Engineering  
 Technical University of Crete  
 Greece, GR73100  
 aggelos@intelligence.tuc.gr

**Georgios Chalkiadakis**

Electronic and Computer Engineering  
 Technical University of Crete  
 Greece, GR73100  
 gehalk@intelligence.tuc.gr

## Abstract

Tackling the decision-making problem faced by a prosumer (i.e., a producer that is simultaneously a consumer) when selling and buying energy in the emerging smart electricity grid, is of utmost importance for the economic profitability of such a business entity. In this work, we model, for the first time, this problem as a factored Markov Decision Process. By so doing, we are able to represent the problem compactly, and provide an exact optimal solution via dynamic programming—withstanding its large size. Our model successfully captures the main aspects of the business decisions of a prosumer corresponding to a community microgrid of *any* size. Moreover, it includes appropriate sub-models for prosumer production and consumption prediction. Experimental simulations verify the effectiveness of our approach; and show that our exact value iteration solution matches that of a state-of-the-art method for stochastic planning in very large environments, while outperforming it in terms of computation time.

## 1 Introduction

In recent years, the term *prosumer* has been coined in order to describe an entity that both produces and consumes energy, implying that prosumers possess the ability to play a key role to the stabilization of the electricity network [Asmus, 2010; Ramchurn *et al.*, 2012]. As such, and assuming prosumers are able to adjust their behaviour according to dynamic indicators, their smooth integration into the shaping *Smart Grid* is of critical importance [Rogers *et al.*, 2012]. Viewed as a business entity, a prosumer could correspond to a single residence, a specific industry, or to whole neighborhoods of houses that are served by a dedicated microgrid—which may or may not be connected to the rest of the electricity Grid. Our focus of attention in this paper will be optimizing the business decisions of a micro grid-corresponding prosumer, producing electricity from (mainly) renewable energy resources, and which has the option of buying and selling energy from utility companies residing in the larger electricity Grid. Paradigms of such community-oriented and renewable energy-relying microgrids will be commonplace in the future [Asmus, 2010].

Notwithstanding its importance, essentially no work to date has, to the best of our knowledge, attacked this specific problem head on. By contrast, in our work, first presented in [Angelidakis and Chalkiadakis, 2015a], we model, for the first time, the decision problem faced by a microgrid-prosumer planning its energy production, storage and usage strategy for the day ahead as a *factored Markov Decision Process* [Boutilier *et al.*, 1999]. Our formulation enables us to provide *an exact optimal solution* (bar certain discretization-related modeling decisions) for the problem faced by a prosumer corresponding to a microgrid of essentially *any* size. In addition, we equip our consumer with specific production-consumption predicting submodels (encompassing Gaussian Processes and Bayesian Linear Regression), which provide the necessary input signals on which to base its decisions.

Given our model, the solution to the prosumer decision problem can then be computed using standard dynamic programming techniques. Specifically, we employed value iteration for this purpose. The effectiveness and efficiency of our approach is verified by comparisons to the performance of *Stochastic Planning using Decision Diagrams (SPUDD)*, a state-of-the-art method for stochastic planning in large environments. Our value iteration method, operating over a problem horizon corresponding to 24 hours, is shown to produce policies that coincide with those produced by SPUDD [Hoey *et al.*, 1999]. However, SPUDD cannot produce a solution in within the required 24-hour timeframe.

The rest of this paper is organized as follows. Section 2 provides a brief background on factored MDPs and reviews related work; Section 3 then describes our model, while the value iteration algorithm is described in Section 4; Section 5 presents our methods for predicting the future production and consumption of the prosumer; Section 6 presents our simulation experiments; and, finally, Section 7 concludes.

## 2 Background and Related Work

*Factored Markov Decision Processes (FMDPs)* [Boutilier *et al.*, 1999] provide a *compact* alternative to standard MDP representation. Specifically, they decompose states into sets of *state variables* in order to represent the transition and model compactly—since transitions and rewards may rely on specific model aspects, corresponding to subsets of variables only. Thus, the set of states in a factored MDP representation correspond to multivariate random variables,  $\mathbf{s} = \langle s_i \rangle$ ,

with the  $s_i$  variables taking on values in their corresponding  $DOM(s_i)$  domains. Intuitively, state variables correspond to a selection of *features* which are sufficient to describe the system state. In FMDPs, actions are also quite often described as random variables, while reward functions used are assumed to be factored into specific (usually additive) components. Furthermore, FMDP models allow for *external signals*, described by *signal variables*, affecting state variables; while *temporal Bayesian networks (TBNs)* and *influence diagrams* can be employed to represent the effects of actions on state transitions and rewards. A multitude of techniques that exploit the resulting representational structure can then be used to solve large problems, at least approximately (e.g., approximate linear programming, stochastic algebraic decision diagrams, and so on) [Guestrin *et al.*, 2003; Boutilier *et al.*, 1999].

Stochastic Planning Using Decision Diagrams (SPUDD) [Hoey *et al.*, 1999], in particular, is a well-known algorithm for finding (near-)optimal policies in very large problems represented as factored MDPs. It is essentially a value iteration algorithm that uses algebraic decision diagrams (ADDs) [Bahar *et al.*, 1997] to represent value functions and policies, assuming an ADD input representation of the FMDP. Specifically, SPUDD operates over an input script describing the factored states and actions, and the FMDP transition model and reward function.

Now there have been a few recent papers dealing with optimal decision-making when buying and selling energy in the Smart Grid. However, most of them do not focus on prosumers. For instance, TacTex [Urieli and Stone, 2014] was the champion agent for the 2013 Power Trading Agent Competition (PowerTAC). In PowerTAC, self-interested, autonomous agents corresponding to *brokers* compete with each other with the goal of maximizing profits through energy trading. Similarly to TacTex, the work of Peters *et al.* (2013) also deals with optimising the long-term behaviour of broker agents during retail electricity trading.

We are only aware of two papers that focus on prosumer decision-making. First, Nikovski and Zhang (2010) propose a method for finding the optimal conditional operational schedule for a set of power generators, assuming stochastic electricity demand and stochastic generator output. However, in contrast to our work, they do not tackle the problem of selling or storing the generated power. Second, Kanchev *et al.* (2011) propose an energy management system which could be employed by a prosumer managing photovoltaic generators, storage units, and a gas microturbine. However, they assume a deterministic system, not accounting for uncertainty and errors during the prosumer’s operation time.

### 3 Our Model

The prosumer we consider in this work corresponds to a *microgrid* distributing power to a community. As such, it produces energy by means of *renewable energy sources*, and is responsible for the well-being of *residential consumers*. Moreover, the prosumer has access to *storage devices (batteries)*, which it can use to store energy for future use. Our prosumer is connected to the wider Grid, and it has to take

decisions regarding the amounts of energy to purchase or sell to the Grid at pre-specified intervals during the next day. We assume that the Grid is represented by some utility company that can specify *tariffs* determining the sell and buy prices of electricity, to which the prosumer can subscribe (at any one of the aforementioned time intervals). The tariffs available to prosumers for the day-ahead are announced by the utility company at the beginning of each day. Then, the problem facing the prosumer is taking the right decisions as to which tariff to subscribe to and what amounts of energy to buy, sell, or store at any given interval of the day-ahead—so as to meet demand at a minimum cost and make a profit by selling the electricity to the utilities.

#### 3.1 Factored Representation

We now describe our problem’s factored representation in some length. More detail can be found in [Angelidakis and Chalkiadakis, 2015a]. To begin, the factored states can be described as a multivariate random variable  $s = \langle s_i \rangle$ , where each variable  $s_i$  can take a value in its domain  $DOM(s_i)$ . There are three factored state variables. The first one, *tms*, takes as values the specific time steps at which the prosumer is able to act. Its domain is originally set to  $[1 \dots 24]$  (one time step per hour in the day). The second one, *bat*, corresponds to the amount of energy available in the batteries, and its domain is  $[0 \dots Battery_{max}]$ , with  $Battery_{max}$  corresponding to the maximum capacity of the storage device(s). Finally, *tf* corresponds to the tariff the prosumer has assigned to at the moment, and its domain is the enumerated tariffs that the utility offers during the day. That is,  $DOM(tf) = \{tf_1, \dots, tf_i, \dots, tf_K\}$ , with  $K$  being the number of tariffs available on a specific day. Each  $tf_i$  tariff is characterized by a buying and a selling price, denoted *buying<sub>i</sub>* and *selling<sub>i</sub>* respectively, and communicated to the prosumer via external signals.<sup>1</sup>

Then, actions can be described as a multivariate random variable  $a = \langle a_i \rangle$  where each variable  $a_i$  affects the transition from some factored state to another, and takes a value in its domain  $DOM(a_i)$ . The discretization for each  $DOM(a_i)$  is performed dynamically: it is based on the discretization of the  $DOM(s_i)$  domains, in a way that from any given state, actions can lead to any other.

There are three factored actions. First, action *buy*, which describes the amount of energy bought from the electric utility. Positive values for *buy* denote the actual buying of energy from the utility, while negative values mean the prosumer sells energy to the utility. With  $Load_{max}$  being the maximum total expected residential consumption load, and the nominal power generating capacity of the renewable energy sources denoted by  $RES_{nom}$ , the domain for *buy* is set to  $[-RES_{nom}$

<sup>1</sup>Notice that tariffs can be key to group together a range of consumer preferences, that would have had to be represented by distinct state or action variables otherwise. For instance, one would have wished to represent preferences to consume when buying prices are low, e.g. at night, and sell when selling prices are high—and distinct sell and buy variables would have been required to allow this. Tariffs could potentially incorporate more information, such as special discounts, and so on. Thus, the use of tariffs can be key at reducing the state-action space in such problems.

...  $Load_{max}$ ]. Second, factored action  $chg$ , which signifies the attempt to store an amount of energy to the batteries. Its value range is  $[-Battery_{max} \dots Battery_{max}]$ . Positive values represent charging the battery, and negative values represent discharging the battery. Finally, the third action,  $sel_{tf}$ , corresponds to a selection of tariff by the prosumer. Its domain is  $[0 \dots K]$ . The value 0 signifies a choice to remain attached to its current tariff, while values 1 to  $K$  signify a choice to move to some other of the  $K$  tariffs available.<sup>2</sup>

Now, there are three types of external signals the prosumer receives. and can be described as multivariate random variable  $sg = \langle signal_i \rangle$  where each variable  $signal_i$  can take a value in its domain  $DOM(signal_i)$ . The prosumer employs these signals in order to determine its actions. The first two,  $prod$  and  $cons$ , specify the current estimates about the production and consumption levels of the prosumer (at a specific time step). Their domains are defined given the  $RES_{nom}$  and  $Load_{max}$  values introduced above. The third signal type,  $price_{tf}$ , specifies, once a day, the buy and sell prices ( $buying_i$  and  $selling_i$ ) for each one of the  $K$  tariffs, and for each  $t$  time step of the day ahead.

Notice that all factored variables in our formulation are independent of the size of the prosumer microgrid—i.e., they are not affected by the number of generators or homes populating it. In what follows, we use the  $x_t$  to denote the value of a state, action, or signal variable at time  $t$ .

### 3.2 Physical Constraints and Transition Function

There are certain constraints that our state and action variables must adhere to. First, in a setting involving energy exchanges, the *balance energy constraint* [Kirschen and Strbac, 2005] must be respected at all times. This means that, at any time step  $t$ , power produced (including that bought) should match power consumed (including that stored):

$$prod_t - cons_t - chg_t + buy_t = 0 \quad (1)$$

The second constraint refers to the storage unit(s) of the prosumer. A storage unit cannot be charged over its capacity:

$$chg_t \leq Battery_{max} - bat_t \quad (2)$$

Similarly, the energy quantity discharged from a unit cannot exceed that currently stored in the unit:

$$-chg_t \leq bat_t \quad (3)$$

Finally, for safety reasons, the battery storage level must be always kept between 20% and 100% [Chiasson and Vairamohan, 2005]:

$$0.2 \leq bat_t / Battery_{max} \leq 1 \quad (4)$$

State transitions in our model are in general stochastic, since faults may occur while taking actions like charging or discharging the storage devices and buying or selling energy to the utility. We define certain *bounded regions* (with distinct boundaries for each variable), which include a subset of discrete factored states lying close to the factored state intended

<sup>2</sup>The additional ‘stay-with-current-tariff’ action is required as subscribing and resubscribing would entail a subscription cost (thus the action protects the prosumer from that cost).

to transition to by performing a factored action taken at time  $t$  (see [Angelidakis and Chalkiadakis, 2015a] for details). The boundaries can be set to any values required. Since distinct factored actions can be simultaneously utilized—i.e., the prosumer can select a new tariff, buy energy, and charge the battery at the same time step  $t$ —the overall transition probability is given by Eq. 5 as follows.

$$Pr(tms_{t+1}, bat_{t+1}, tf_{t+1} | tms_t, bat_t, tf_t, chg_t, sel_{tf,t}) = Pr(bat_{t+1} | bat_t, chg_t) \cdot Pr(tf_{t+1} | tf_t, sel_{tf,t}) \quad (5)$$

given that the battery level at any time step depends on the previous battery level state and on whether a  $chg$  action was used, while the tariff in place is affected by tariff selection.

### 3.3 Factored Reward Representation

The next step is to determine the reward function for our factored MDP. The reward function is associated with (a) either the gain from selling power to the utility or the cost of buying power in a certain price; (b) the running costs for being subscribed to a tariff; and (c) the operation costs of using the storage devices. As such, we choose to represent the reward function as a cost function with three main components. Specifically, the function describing the immediate cost for a transition from state  $s_t$  to  $s'_{t+1}$  by executing some  $a_t$  at time-step  $t$ , is defined as follows:

$$Cost(s_t, a_t, s'_{t+1}) = C_{energy} + C_{period} + C_{bl} \quad (6)$$

The first component,  $C_{energy}$ , captures the cost per Wh for buying electricity (or the profits from selling it to the utility), given the buy/sell rates prescribed by the tariff in effect. The second component captures the periodic costs  $C_{periodic}$  inflicted on the prosumer for being subscribed into a tariff: The third component of the cost function,  $C_{bl}$ , captures the costs associated with *battery life losses*. That is, the costs inflicted from charging (or discharging) the storage devices (batteries) with a charge amount of  $chg_t$ , at a given time-step  $t$  when the stored energy amount is at  $bat_t$ .

## 4 Solving the FMDP

With the above FMDP at hand, the optimal policy can be derived by solving the corresponding Bellman equations. Dynamic programming (DP) methods can be used to obtain the optimal solution [Sutton and Barto, 1998]. In our work here we used *value iteration (VI)* as the DP method of choice. Our problem is naturally a finite-horizon problem, thus we employed a finite-horizon VI method. By setting the horizon  $T$  to be equal to the number of time steps at which the prosumer is required to act, we can incorporate the  $tms$  factored state into the problem’s horizon, thus effectively reducing the size of the state space.

Then, with  $s'_t$  denoting the potential successor states of  $s_t$ ; with  $Pr(s'_{t+1} | a_t, s_t)$  denoting the probability of state transitions from  $s_t$  to possible successor states  $s'_{t+1}$ , given that action  $a_t$  was taken; and  $R(s_t, a_t, s'_{t+1}) = -Cost(s_t, a_t, s'_{t+1})$  denoting the corresponding immediate reward (the negative immediate cost), the VI algorithm iteratively estimates the value function for the factored states, and outputs an optimal policy  $\pi$ , as shown in Alg. 1.

```

for all instantiations of  $s$  do
  | set  $V_{T+1}(s) = 0$ 
end
for  $t$  stages-to-go (i.e., with  $1, \dots, T$  stages-to-go) do
  | for all instantiations of  $s_t$  do
    |  $V_t(s_t) \leftarrow \max_{\mathbf{a}_t} \sum_{s'_{t+1}} Pr(s'_{t+1} | \mathbf{a}_t, s_t) \cdot$ 
    |  $(R(s_t, \mathbf{a}_t, s'_{t+1}) + V_{t+1}(s'_{t+1}))$ 
    end
  end
for all instantiations of  $s$  and all  $t$  stages-to-go do
  |  $\pi(s, t) =$ 
  |  $\arg \max_{\mathbf{a}} \sum_{s'} Pr(s' | \mathbf{a}, s) (R(s, \mathbf{a}, s') + V_{t+1}(s'))$ 
end

```

**Algorithm 1:** Value iteration for solving the FMDDP

## 5 Production and Consumption Models

Naturally, the estimated production from the renewable energy sources distributed on the microgrid, and the predicted load consumption of the connected consumers, affect the policy of the prosumer. The prosumer is notified about the expected production and consumption values via the *prod* and *cons* signals. Thus, it is necessary to predict values for those signals that are as accurate as possible, to assist the decision-making process of the prosumer.

To obtain the production estimates of the photovoltaic systems (PVS) and wind turbine generators (WTG) of our microgrid, we employ *RENES* [Panagopoulos *et al.*, 2012], a web-based PVS and WTG production prediction tool. *RENES* generates PVS and WTG production estimates given time, geographical coordinates and online weather forecasts, and it comes with specific performance guarantees.

Then, in order to predict the load consumption of the prosumer we evaluated the use of Gaussian Processes (GPs) and Bayesian Linear Regression [Bishop, 2006]. Our evaluation demonstrated that GP with Gaussian kernel performs better and is used for consumption prediction in our work.

## 6 Experiments and Results

We evaluate our model by examining a residential prosumer at New Hampshire, New England, northeastern United States. The data used in our prediction of residential load consumption for the area, comes from the Public Service Company of New Hampshire, and is freely available in their website (<http://www.psnh.com/>). Our simulated prosumer serves 30 households and includes 20 photovoltaic modules with nominal power 60kW, 2 windturbines with nominal power 1000kW and 24 deep cycle 12Volts batteries, with cost of each battery 269,00 €. Estimated battery lifetime is 10-12 years. All experiments were conducted on a 2.10 GHz x 4 Intel Core i3-2310M processor, with 8GB of memory.

Our results, reported in [Angelidakis and Chalkiadakis, 2015a], show that the optimal policy computed through value iteration and SPUDD for the day-ahead coincide with each other, for state-action spaces with various sizes. Nevertheless, value iteration produced the optimal policy in approxi-

Horizon	$ S \times A $	Bounded Region Size	Value Iteration (hours)	SPUDD (hours)	
				Script	Runtime
24	664290	15	1.76	13.4992	0.184
		90	15.84	46.9188	1.19
	2624490	15	8.7603	36.98	0.73975
48	664290	15	3.5	16.8221	0.4271

Table 1: Running time of value iteration and SPUDD for four different scenarios. “Script” refers to the pre-processing time required for the SPUDD input files to be generated, while “Runtime” denotes the subsequent SPUDD execution time.

mately 15% or 25% of the required time for SPUDD to extract the same policy. Moreover, SPUDD was often not able to produce a solution within 24 hours, and could not generate a final policy with the available memory, in contrast to our VI method. The exact running times are presented in Table 1.

Our experiments thus demonstrate the limitations of SPUDD when used for problems that do not possess enough structure to allow for a compact enough representation of the required transitions in its input files.

## 7 Conclusions

This paper employs, for the first time, factored MDPs to model the decision problem faced by a prosumer planning its energy flow management for the day-ahead. Our model incorporates the key factors responsible for the effective operation of a microgrid prosumer, regardless of its size; and allows us to obtain the exact optimal solution to the problem. We used a simple value iteration algorithm to compute the solution to this sequential decision making problem, and demonstrated our method’s effectiveness and efficiency by comparing it to the performance of *SPUDD*. By so doing, we exposed the limitations of this particular FMDDP solver. While our model enables the simple VI method to compute the optimal solution within a reasonable time, the problem does not have enough structure to allow the creation of a compact input file for SPUDD to operate on, resulting to poor performance. In addition, this work provides specific predictive tools for obtaining prosumer consumption and production estimates, and exhibits how *Gaussian processes* and *Bayesian linear regression* techniques can be used for consumption prediction in this setting (with GPs emerging as the most successful).

In a recent follow-up paper to this work [Angelidakis and Chalkiadakis, 2015b], we used approximate MDP solution methods for taking decisions in this domain without the need of discretizing the state space. Specifically, we employed fitted value iteration, a sampling-based approximation method that is known to be well behaved [Munos and Szepesvári, 2008]. By so doing, we generalized our factored MDP solution method to continuous state spaces. Future work includes more prosumer actions into our model and incorporating it within *renewable energy sources cooperatives*<sup>3</sup>, the emergence of which is of extremely high economic, social, and environmental importance [Chalkiadakis *et al.*, 2011].

<sup>3</sup> <http://www.rescoop.eu>.

## References

- [Angelidakis and Chalkiadakis, 2015a] A. Angelidakis and G. Chalkiadakis. Factored mdps for optimal prosumer decision-making. In *Proc. of the conference on Autonomous Agents and Multi-Agent Systems (AAMAS) - 2015*, pages 503–511, 2015.
- [Angelidakis and Chalkiadakis, 2015b] A. Angelidakis and G. Chalkiadakis. Factored mdps for optimal prosumer decision-making in continuous state spaces. In *Proc. of EUMAS-2015*, 2015.
- [Asmus, 2010] P. Asmus. Microgrids, virtual power plants and our distributed energy future. pages 72–82, 2010.
- [Bahar et al., 1997] R. Bahar, E. Frohm, C. Gaona, G. Hachtel, E. Macii, A. Pardo, and F. Somenzi. Algebraic decision diagrams and their applications. pages 171–206, 1997.
- [Bishop, 2006] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [Boutilier et al., 1999] C. Boutilier, T. Dean, and S. Hanks. Decision-theoretic planning: Structural assumptions and computational leverage. pages 1–94, 1999.
- [Chalkiadakis et al., 2011] G. Chalkiadakis, V. Robu, R. Kota, A. Rogers, and N. Jennings. Cooperatives of distributed energy resources for efficient virtual power plants. In *Proc. of the conference on Autonomous Agents and Multi-Agent Systems (AAMAS) - 2011*, pages 787–794, 2011.
- [Chiasson and Vairamohan, 2005] J. Chiasson and B. Vairamohan. Estimating the state of charge of a battery. pages 465–470, 2005.
- [Guestrin et al., 2003] C. Guestrin, D. Koller, R. Parr, and S. Venkataraman. Efficient solution algorithms for factored mdps. pages 399–468, 2003.
- [Hoey et al., 1999] J. Hoey, R. St-Aubin, A. J. Hu, and C. Boutilier. Spudd: Stochastic planning using decision diagrams. In *Proc. of the conference on Uncertainty in Artificial Intelligence (UAI) - 1999*, pages 279–288, 1999.
- [Kanchev et al., 2011] H. Kanchev, F. Colas, D. Lu, V. Lazarov, and B. Francois. Energy management and operational planning of a microgrid with a pv-based active generator for smart grid applications. pages 4583–4592, 2011.
- [Kirschen and Strbac, 2005] D. Kirschen and G. Strbac. *Fundamentals of Power System Economics*. J. Wiley & Sons, 2005.
- [Munos and Szepesvári, 2008] R. Munos and C. Szepesvári. Finite-time bounds for fitted value iteration. In *Proc. of Journal of Machine Learning Research*, pages 815–857, 2008.
- [Nikovski and Zhang, 2010] D. Nikovski and W. Zhang. Factored markov decision process models for stochastic unit commitment. pages 28–35, 2010.
- [Panagopoulos et al., 2012] A. Panagopoulos, G. Chalkiadakis, and E. Koutroulis. Predicting the power output of distributed renewable energy resources within a broad geographical region. In *Proc. of European Conference on Artificial Intelligence (ECAI) / Prestigious Applications of Intelligent Systems (PAIS) - 2012*, pages 981–986, 2012.
- [Peters et al., 2013] M. Peters, W. Ketter, M. Saar-Tsechansky, and J. Collins. A reinforcement learning approach to autonomous decision-making in smart electricity markets. pages 5–39, 2013.
- [Ramchurn et al., 2012] S. D. Ramchurn, P. Vytelingum, A. Rogers, and N. Jennings. Putting the 'smarts' into the smart grid: A grand challenge for artificial intelligence. pages 86–97, 2012.
- [Rogers et al., 2012] A. Rogers, S. Ramchurn, and N. Jennings. Delivering the smart grid: Challenges for autonomous agents and multi-agent systems research. pages 2166–2172, 2012.
- [Sutton and Barto, 1998] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 1998.
- [Urieli and Stone, 2014] D. Urieli and P. Stone. Tactex'13: a champion adaptive power trading agent. In *Proc. of the conference on Autonomous Agents and Multi-Agent Systems (AAMAS) - 2014*, pages 1447–1448, 2014.