IMT Institute for Advanced Studies, Lucca

Lucca, Italy

Prosumer planning in the DEZENT context of regenerative power production

PhD Program in Computer Science and Engineering

XXV Cycle

By

Alain Tcheukam Siwe

2013

The dissertation of Alain Tcheukam Siwe is approved.

Program Coordinator: Prof. Rocco De Nicola, IMT Institute for Advanced Studies Lucca, Italy

Supervisor: Prof. Ugo Montanari, University of Pisa, Italy

Tutor: Prof. Ugo Montanari, University of Pisa, Italy

The dissertation of Alain Tcheukam Siwe has been reviewed by:

Prof. Haralambos Sarimveis, NTUA Athens, Greece

Prof. Horst Wedde, University of Dortmund, Germany

IMT Institute for Advanced Studies, Lucca

2013

To my father Thomas Tcheukam.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisor Prof. Ugo Montanari (Department of computer science, University of Pisa) for the advices and supports provided for carrying out this work. I also thank professor Horst Wedde from the School of Computer Science (TU Dortmund, Germany) for the research activity carried out within the research group DEZENT. Finally, I acknowledge support by the EU Integrated Project 257414 ASCENS, and by the Italian MIUR Project CINA (PRIN 2010).

Last, but not least, I would like to thank my family, my wife and her family.

Vita

July, 2013	Winner of the award -Energy for knowledge- supported by Fondazione Centro Studi Enel and Fondazione CRUI, Italy.
April, 2013	Research Fellowship, Department of Computer Science, University of Pisa, Italy.
March, 2010	PhD student in Computer Science and Engineering, XXV Cycle Ph.D. Program at IMT- Institute for Advanced Studies at Lucca, Italy.
October, 2009	M.S. degree with summa cum laude in computer engineering from the University of Padua, Italy.
2008	Selected for an annual exchange position according to the agreement between the University of Padua and the Illinois Institute of Technology of Chicago (USA).
November, 2007	B.S. degree in computer engineering (100/110) from the University of Padua, Italy.
2004	Qualification as a machine tool operator - ENAIP VENETO of Longarone (BL), Italy.
2002	Diploma of General Academic Study in Mathematics from the Normal School of the University of Yaound 1, Cameroon. French scholastic program.
January 22, 1983	Born in Cameroon.

Publications

1. Ugo Montanari and Alain Tcheukam. Real time market models and prosumer profiling. *IEEE Conference on Computer Communication Workshop*, pages 7-12, April 2013. Turin, Italy.

Presentations

1. Alain Tcheukam, session theme: Technologies and Strategies for Demand Response, *Energy for knowledge meeting - Center of Study Enel* , july 11, 2013, Rome, Italy.

2. Ugo Montanari and Alain Tcheukam. Real time market models and prosumer profiling. *IEEE Conference on Computer Communication Workshop*, pages 7-12, April 2013. Turin, Italy.

3. Alain Tcheukam, Formal Market Models for Smart Power Grids, *Invited talk, University of Modena and Reggio Emilia*, February 16, 2012, Italy.

Abstract

The electricity is a vital asset and a priority for the social and economic development of today's world. Building energy infrastructures with high efficiency and renewable energy sources is an important yet challenging task for a sustainable future. Smart grid is a term referring to a modernized electrical grid that uses information and communications technology to gather and act on information, such as information about the behaviors of suppliers and consumers, in an automated fashion to improve the efficiency, reliability, economics, and sustainability of the production and distribution of electricity.

The main issue of this thesis is to propose new solutions which help end-users to optimize their consumption and better manage their own electricity costs. More specifically, the challenge is to make elastic the demand for, and the supply of, electricity of *prosumers* in order to optimize their energy cost based on power market conditions and on suitable constraints on their power consumption. By definition, a prosumer is a user that not only consumes, but also produces and stores electricity.

In our work, we focus on power market models in which prosumers interact in a distributed environment during the purchase or sale of electric power. We have chosen to follow the distributed power market model DEZENT. Our contribution is the planning phase of the consumption of prosumers based on the negotiation mechanism of DEZENT. We propose a controller for the planning of the consumption which aims at minimizing the electricity cost achieved at the end of

a day. Our controller model exploits the standard dynamic programming algorithm and in the thesis we discuss the assumptions on which the controller design is based.

In order to evaluate the performance of our introduced controller, we performed extended experimental studies based on the available DEZENT simulator and on the Java implementation of the optimal controller. Our main result is that the highest energy cost reduction was obtained when we have a high variance on the profile cost of the electricity, the prosumer environment is in the undersupply situation and the reserve capacity of the controller is infinite. Vice versa, the lowest energy cost reduction corresponds to low variance, oversupply situation and finite reserve capacity. Furthermore the study of the problem of (sub) optimal repeated re-planning for the rest of the day has shown that a prosumer having to consume more than expected will pay a remarkable additional cost at the end of the day which depends also on the increased unit costs; in the case in which a part of the available energy reserve is lost, the additional cost paid is proportional to the amount of energy lost.

In summary the general idea behind the planning (optimization) phase of a consumer is to plan consumption as smartly (delay or anticipate the consumption) as possible during a day, a week, a month or even a year. The results of this behavior is the minimization of the electricity cost in the long run, under certain assumptions on energy costs, as resulted by local negotiations, and on the acceptable variations of consumer requests. One of the open issues is the global effect of our introduced controller in a prosumer population since each prosumer can make use of it. We believe that the issue can lead to a congestion problem similar to that of the minority game problem proposed in economics literature.

Chapter 1

Introduction

1.1 Presentation of the area

As global perspective, Smart Grid means different things in different countries (China, Brazil, Singapore, India and USA) [1, 2, 3, 4]. Although there is no standard global definition, the European Technology Platform Smart Grid defines smart grids as electricity networks that can intelligently integrate the behavior and actions of all users connected to it such as producers, consumers and prosumer[1]. The aim is to efficiently deliver sustainable, economic and secure electricity supplies through the use of innovative products and services together with intelligent monitoring, control, communication, and self-healing technologies [5]. The power grid operation can be subdivided into three main task (see Figure 1):

- *Centralized power generation*: where electricity is produced by large size energy source generation as coal plant, nuclear plant, natural gas plant and hydroelectric plant. Generally, the electricity is produced in a extra high voltage in the order of 265 to 275 KV.

- *Power transmission*: it is a high voltage electric transmission in the order of 110 KV or above, from generating power plants to sub-

[1]Prosumer: a user that not only consumes electricity, but also produces and stores electricity.

stations located near to population centers. The electricity is transmitted at high voltages to reduce the energy lost in long distance transmission and power is usually transmitted through overhead power lines.

- *Electricity distribution*: it is the final stage in the delivery of electricity to end users or consumers. It works typically at medium-voltage (less than 50 KV) and low-voltage (less than 1 KV).

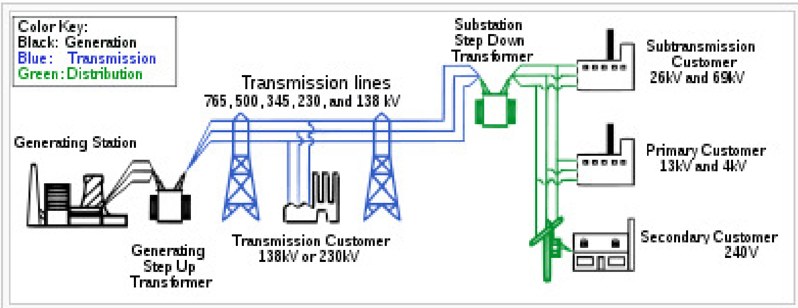

Figure 1: Diagram of an electric power system

Furthermore the power grid sector is subject to limitations which affect the market model organization. We can refer to network externalities which are due to technological limits: the electricity cannot be infinitely stored and must therefore be generated as it is needed; and power line losses are due to long distance transmission (this is a more general problem and we do not handle it). These limitations generate a problem of stability in the management of the electricity network and the need to ensure a balance between the electricity production and consumption [6]. In fact, in the worst case scenario the difference between the electricity supply and the electricity demand can be very large. This can lead to blackout situations (under production) or storage issues (over production). Furthermore the mechanism used to guarantee the close match between load and generation is coordinated through the electricity market. In a regulated market, broker companies offer fixed prices to end consumers independently of the demand and supply situation, and most of the time users have no way to understand the pricing mechanisms which affect the growing cost of the electricity.

The continuously increasing cost of the electricity along with the need to reduce greenhouse gas emissions to protect the environment, have made energy efficiency one of the technological challenges of our century. The purpose is to optimize the grid operation and the electricity usage worldwide. We can refer to the EU's 20-20-20 climate change objectives, whose target for the year 2020 includes: 20% reduction in greenhouse gas emissions, 20% EU renewables share and 20% savings in consumption by improving energy efficiency [7]. These goals require new solutions and management strategies at the (see Figure 2):

1. power and energy layer;
2. communications layer;
3. computing and information technology layer.

In the power and energy layer there is a need to integrate renewable energy sources in the electricity production chain, such as solar and wind, in order to reduce the peak of energy consumption and transmission losses. This may lead to a decentralization of the power grid due to the electricity production from the medium and low voltage layers of the grid; and to the need to take into account the intermittence of renewable energy sources during the electricity production. In the communication layer, and in the computing and information technology layer, both electric power and information flows will be distributed. The need to reduce energy demand imbalances will require tackling the optimization of the electricity generation, transportation and distribution. This will be possible by controlling the real-time collection of data on the status of systems and on the network, and by using advanced technologies of communication and elaboration of data based on models of distributed computing and on adaptive algorithms.

Consumer side management will change and new scenarios will occur. For instance, when power is least expensive the user can turn on selected home appliances such as washing machines or factory processes

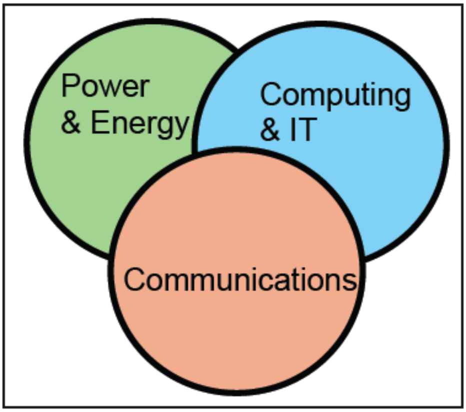

Figure 2: Smart grid foundational layers

that can run at arbitrary hours, while at peak time they could turn off selected appliances to reduce demand or shift usage. End-users will now plays different roles acting as producer or consumer and thus they will contribute to energy saving in the network. As a result, also the electricity market place may change and become open and competitive (standard economic viewpoint). Consumers may be capable to express preference according to tariff price or contract offers by choosing the best company or broker agent. Market decentralization will lead to both big and small player participation in the electricity market scenario and consequently competition will increase. The electricity market may have to change its operation according to the stakeholder (producer, consumer, and broker). In addition, the entrance of plug-in electric vehicles (PEV) will increase the load factor of the grid. PEV could buffer renewable power sources such as wind or solar power, for example, by storing excess energy produced during windy or sunny periods and returning it back to the grid during high load periods, thus effectively stabilizing the intermittency of wind or solar power.

The electricity challenges mentioned above will require heavy investments, and someone would have to pay for their costs. Will it be the state or the power companies? Because of capital exposure and risk man-

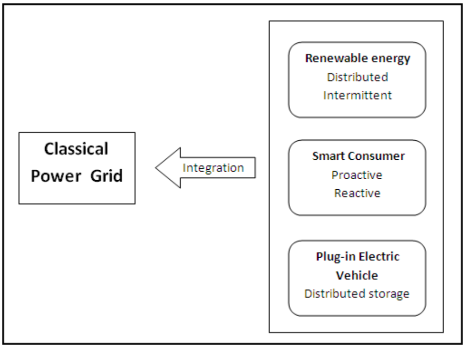

Figure 3: Components to be integrated in the traditional grid system

agement, a shift of interest of investors (state or companies) from large scale power generation plants to medium and low generation renewable energy sources is another challenge to be addressed. In European countries, governmental aims to increase the portion of sustainable energy in the national energy mix have been translated into incentives and tax policies to promote the uptake of renewable energy sources. Reduced energy price would be undoubtedly one of the key results of the challenge of transforming the energy system to become sustainable. Despite these efforts and as a result of the deregulation of power grid, the long-term prospects for large-scale investments in power generation are unclear at this moment. The investments will be possible only if the impact of the reconstruction will be to increase the efficiency of power grid management.

1.2 General description of work

Decentralized power management systems will play a key role in reducing greenhouse gas emissions and increasing electricity production through alternative energy sources. In this thesis, we focus on power market models in which prosumers interact in a distributed environment during the purchase or sale of electric power. We have chosen to follow the distributed power market model DEZENT [49, 50]. Our contribution [42] is the planning phase of the consumption of prosumers based on the negotiation mechanism of DEZENT. We propose a controller for the planning of the consumption which aims at minimizing the electricity cost achieved at the end of a day. In the thesis we discuss the assumptions on which the controller design is based.

Motivation

The continuously increasing cost of the electricity along with the need to reduce greenhouse gas emissions to protect the environment, have made energy efficiency one of the technological challenges of our century. In particular, one of the challenges of the smart power grids is to make elastic the demand/offer of electricity of prosumers connected to the grid. A prosumer (producer-consumer) is defined as a user that not only consumes, but also produces and stores electricity. An improved integration of renewable energy production in the power grid for a better balancing of energy will require both negotiation mechanisms which favor local access and prosumer planning policies which take into account the cost of energy at different times (of day, week, year).

Background power markets

In the new era of smart power grid, we can distinguish between two types of electricity power management systems: the centralized [19, 46] and the decentralized [49] system. The centralized power management system is currently used in many countries. The main feature of this model is that at the physical layer, the grid is designed for a one-way flow of the electricity. More precisely energy flows from the top (where the electricity is generated in large power plants and transported to local substations) to the bottom (final stage in the delivery of electricity to end users). The wholesale power market (see Robert Wilson [54]) can be subdivided into two categories: integrated (or pool) market and unbundled (or forward) market.

DEZENT

The idea behind a decentralized power management system is to exploit the increasing integration of decentralized energy resources (DER) into the distribution network. DER systems are exploiting modern technologies based on solar or wind power, or on other renewable energy sources. In particular we refer to the real-time and distributed power management system DEZENT. DEZENT [49, 50] is the result of an R&D project involving the School of Computer Science and the College of Electrical Engineering of the Dortmund University, the E.ON Energy company and DFG, the German Research Foundation. The project was devoted to decentralized and adaptive electric power management through a distributed real-time multi-agent architecture. DEZENT philosophy is to make as local as possible the exchange of electricity between various prosumers. Since the power market is managed in a decentralized manner, prosumers able to carry out more exchanges at the local level will get more benefits.

In the DEZENT initiative, a multi-level bottom-up solution has been developed where autonomous collaborative software agents negotiate available energy quantities (the DEZENT algorithm [49]). Moreover, since much of the production (e.g. from wind farms or from solar panels) is highly unpredictable, the distributed negotiations algorithm of DEZENT is finalized within short intervals of 0.5 sec. This approach guarantees that the negotiation process is stable, or constant, enough, for a fair and transparent bidding and offering of the partners involved. During each slot of 0.5 sec, the negotiation algorithm proceeds bottom-up, from the 0.4 KV to the 110 KV layers, guaranteeing the balancing of electric power. In addition, a Demand Side and Supply Management (DSSM) algorithm [51] is applied in order to reduce the need of negotiating in the free power market. At the end of the negotiation, each prosumer independently and simultaneously adjusts his/her bidding strategies. This is carried out by the DECOLEARN algorithm [52] which exploits reinforcement learning principles of machine learning.

Problem definition

A smart grid is an electricity network that can intelligently integrate the actions of all prosumers connected to it, in order to efficiently deliver sustainable, economic and secure electricity supplies. The main issue is to propose new solutions which help end-users to optimize their consumption and better manage their own electricity costs. More specifically, the challenge is to make elastic the demand for, and the supply of, electricity of prosumers in order to optimize their energy cost based on power market conditions and on suitable constraints on their power consumption.

Contribution

Our contribution is the planning phase of the consumption of prosumers based on the negotiation mechanism of DEZENT. We believe that independent planning by the prosumers may improve significantly the matching between production and consumption in the DEZENT power grid. In fact, each prosumer should try as much as possible to independently modify his/her power requirements optimizing his cost. In practice, this could mean to help balancing the power market, since the price will favor low consumption/high production when the cost is high and vice versa. Moreover, our approach is not centralized. The idea is to exploit the (limited) ability of prosumers of planning their consumption/production. Hence they do not sign any contract leaving the planning to others: our independent consumer planning is a local matter involving only one prosumer. The planning phase [6] we propose is based on optimization techniques and exploits an efficient dynamic programming algorithm.

Furthermore we study the problem of (sub) optimal repeated re-planning for the rest of the day (of an active consumer) when the consumption varies wrt. the planned profile. These changes can be due to the need of additional power - which however does not require any changes of the planned profile - or to the loss of a part of the available energy reserve. Finally we study the effect of the capacity of the energy reserve on the prosumer ability of planning their consumption.

Methodology

The main issue is that electricity may have a rather different price depending on the time period, be it of the day, of the week or of the year. More specifically, it depends on several exogenous factors, e.g. on the actual cost of energy at the global level, but also on the existence of convenient energy exchanges at the local level and on the fairness of available market mechanisms. The case we consider is that of a decentralized power market like DEZENT, where prices can change rapidly and users may not have explicit information about electricity cost. Hence the behavior (consumption) of the consumer will be based on the information (s)he has on (i) the estimated cost of the electricity and (ii) the consumption constraints. His goal is to minimize the electricity cost at the end of a day while consuming the same amount of energy. To simplify our study, we make some assumptions, which highlight certain aspects of the problem.

1. The DEZENT negotiation process and the DECOLEARN algorithm prevent agent selfishness and their convergence does not heavily depend on the cost of energy or on the amount of power actually exchanged, which can vary a lot e.g. during the day. Those parameters are just scale factors.

2. From the point of view of the prosumer, the unit cost of energy after the DEZENT negotiation depends mainly on two factors: (a) the free market power cost; (b) the prosumer environment: heavy production, heavy consumption or equilibrated prosumer population may yield rather different energy costs. The variance of the free market power cost is particularly relevant: high variance does offer the prosumer a better chance of allocating his/her power needs when it is more convenient economically.

3. The amount of gain depends on the ability of the prosumer at modifying his/her demand and supply needs. We consider a day-long profile and divide it in a number of slots. Moreover, we assume that a prosumer has the ability of increasing or decreasing, up to some

amount, the required power. In addition, he/she has a bounded reserve: thus given the prosumer original consumption profile, the sum of all the positive/negative changes should never exceed the available reserve.

4. If the prosumer knew the actual cost of energy in every slot, he/she could compute (by dynamic programming) a profile of changes which, among the allowed profiles, would optimize the gain. Of course he/she cannot foresee the future. A reasonable estimate of the cost of energy in every slot can be obtained by looking at the values in the previous day. In fact, if we assume that the free market cost in the same slot of the previous day is the same, that the prosumer environment is the same and that the DECOLEARN algorithm is close to convergence, we can safely rely on the after-DECOLEARN costs of the previous day. Of course the prosumer environment has a stochastic behavior, thus if might behave differently in the two days. To improve the estimate, we could take an average of several days, or we could install a new reinforcement learning process. However both approaches apparently do not offer significant improvements experimentally, while in addition the latter choice is computationally very expensive, since different slots should be equipped with different parameters.

5. In summary, our approach depends on the following parameters: (a) the profile of the free market power cost in every slot, and in particular its variance; (b) the prosumer environment: heavy production, heavy consumption or equilibrated prosumer population; (c) how to estimate the cost of energy after the DEZENT negotiation; (d) the amount of allowed power increase/decrease; and (e) the available reserve. Items (a) and (b) affect the simulations but their values are not directly available to the prosumer, while items (c), (d) and (e) are available inputs for prosumers optimal planning.

Experimental results

We performed extended experimental studies based on the available DEZENT simulator and on a Java implementation of the optimal controller. An experiment depends essentially on three parameters: (1) the free market power cost, which can exhibit high, low or average variance: for this we chose real data from the day ahead market prices of Switzerland, Italy and Germany respectively; (2) the prosumer environment, namely heavy production, heavy consumption or equilibrated prosumer population; and (3) the available reserve capacity which can be either finite or infinite. For every combination of the parameters we synthesized the optimal controller and computed the gain with respect to a neutral consumer operating in the same environment (i.e. within a context of other prosumers, which are the same in both cases). As it could be expected, the highest energy cost reduction was obtained when we have a high variance on the profile cost of the electricity, the prosumer environment is in the undersupply situation and the reserve capacity of the controller is infinite. Vice versa, the lowest energy cost reduction corresponds to low variance, oversupply situation and finite reserve capacity. The most contentious item, i.e. how to estimate the cost of energy after the DEZENT negotiation, turned out not to be very relevant. In fact, computing the optimal profile with the power costs of the previous day, or, anticipating the future, of the present day, did not produce large differences in the final cost.

An interesting question is what is the effect on the DEZENT negotiation process of the increased power demand in the low cost periods (and decrease during high cost) caused by our planning approach. We observed a limited negative feedback effect: more demand caused an increase in cost, which in turn made less convenient to increase power demand in that period. We measured this effect by comparing the unit costs of the neutral and of the active user after the DEZENT negotiation. However the negative feedback effect is anyway included in our final cost comparisons between the neutral and the active consumer.

The study of the problem of (sub) optimal repeated re-planning for the rest of the day has shown that a prosumer having to consume more than expected will pay a remarkable additional cost at the end of the day which depends also on the increased unit costs; in the case in which a part of the available energy reserve is lost, the additional cost paid is proportional to the amount of energy lost. However, replanning is definitely convenient with respect to continuing with the old plan, just adapted for the loss of power.

Summary

The general idea behind the planning (optimization) phase of a consumer is to plan consumption as smartly (delay or anticipate the consumption) as possible during a day, a week, a month or even a year. The results of this behavior is the minimization of the electricity cost in the long run, under certain assumptions on energy costs, as resulted by local negotiations, and on the acceptable variations of consumer requests. Future work might consider particular kinds of prosumers and their energy storing devices, e.g. batteries of electric cars, modeling their peculiar abilities of adapting their needs to variable energy cost profiles.

1.3 Synopsis

The remainder of this thesis is organized as follows. In chapter 2 theoretical backgrounds on reinforcement learning are considered both in the computer science and economics perspective. Later in chapter 3, smart power grid scenarios are considered. In chapter 4 we give a more detailed description of the power market model of DEZENT from which our study is based. A controller is then proposed in chapter 5 for planning the consumption of a prosumer connected to the grid. The performance of our introduced controller is studied through experiments in chapter 6. Finally in chapter 7 we present conclusions and future works.

Chapter 2

Background on Reinforcement Learning

In this chapter, the theoretical background on reinforcement learning is presented. In particular, we refer to the field of computer science (machine learning), where the basics of reinforcement learning are studied, and to the field of economics and game theory (minority games), where reinforcement learning is seen as a particular game. In a few words, the idea behind reinforcement learning is that there is an agent interacting with a dynamic environment about which it has no information: the agent's goal is to learn the behavior of the environment and then to modify its actions accordingly. Reinforcement learning can be subdivided into 2 steps: the choice of the action to be executed and the reward to be assigned to that action. As the agent has no exact information about the state of the environment (it depends on the behavior of other agents involved), the choice of an action is made probabilistically, according to a distribution which depends on past experiences. The reward assigned to the action taken is based on the result of the interaction with the environment. The probabilities of actions are then modified in such a way to increase the long-run reward.

2.1 Reinforcement learning

2.1.1 Literature on Reinforcement Learning

Behavioral psychology [8], also called learning perspective, is a philosophy of psychology based on the proposition that all things that organisms do, including acting, thinking, and feeling can and should be regarded as behaviors. Inspired by behavioral psychology, reinforcement learning is an area of machine learning within computer science, concerned with how an agent ought to take actions in an environment so as to maximize some notion of cumulative reward. The problem, due to its generality, is studied in many other disciplines, such as game theory, operations research, information theory, simulation-based optimization, statistics, and genetic algorithms. The problem has been studied in optimal control theory, though most studies there are concerned with the existence of optimal solutions and their characterization, and not with the learning or approximation aspects.

In economics and game theory [28], reinforcement learning may be used to explain how equilibrium may arise under bounded rationality. Bounded rationality is the idea that in decision-making, rationality of individuals is limited by the information they have, the cognitive limitations of their minds, and the finite amount of time they have to make a decision. It was proposed by *Herbert Simon* as an alternative basis for the mathematical modeling of decision making, as used in economics and related disciplines. Bounded rationality complements optimization, which views decision-making as a fully rational process of finding an optimal choice given the information available.

In computer science [39], reinforcement learning is the problem faced by an agent that must learn the behavior through trial-and-error interactions with a dynamic environment. There are two main strategies for solving reinforcement-learning problems: the first is to search in the space of behaviors in order to found one that performs well in the environment and this approach has been taken by work in genetic algorithms and genetic programming; the second is to use statistical techniques and dynamic programming methods to estimate the utility of taking actions in states of the world. Hereafter we are focusing on statistical techniques because they take advantage of the special structure of reinforcement-learning problems, which is not available in optimization problems in general [39]. In the standard reinforcement-learning model, an agent is connected to its environment via perception and action, as depicted in figure 4. On each step of interaction the agent receives as input, i, some indication of the current state, s, of the environment; the agent then chooses an action, a, as output. The action changes the state of the environment, and the value of this state transition is communicated to the agent through a scalar reinforcement signal, r. The agent's behavior, B, should choose actions that tend to increase the long-run sum of values of the reinforcement signal. It can learn to do this over time by systematic trial and error, guided by a wide variety of algorithms.

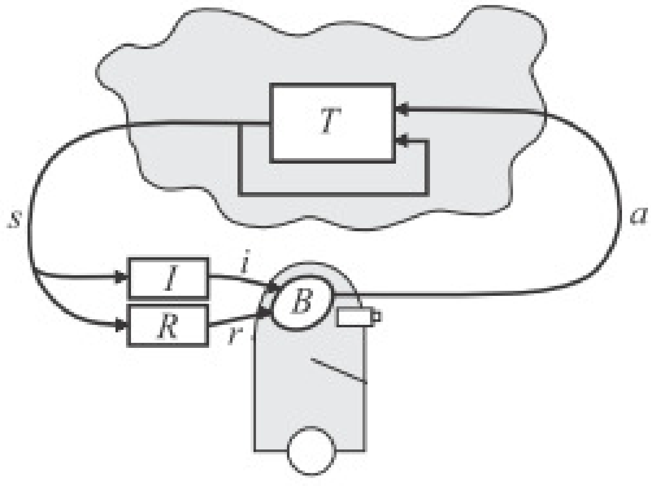

Figure 4: The standard reinforcement-learning model.

For example, the simplest possible reinforcement-learning problem is known as the k-armed bandit problem [26], which has been the subject of a great deal of study in the statistics and applied mathematics literature (Berry & Fristedt, 1985). The agent is in a room with a collection of k

gambling machines (each called a "one-armed bandit" in colloquial English). The agent is permitted a fixed number of pulls, h. Any arm may be pulled on each turn. The machines do not require a deposit to play; the only cost is in wasting a pull playing a suboptimal machine. When arm i is pulled, machine i pays off 1 or 0, according to some underlying probability parameter p_i , where payoffs are independent events and the p_is are unknown. The major issue is: what should the agent's strategy be? The k-armed bandit problem illustrates the fundamental tradeoff between exploitation and exploration. In fact, the agent might believe that a particular arm has a fairly high payoff probability; should it choose that arm all the time, or should it choose another one that it has less information about, but seems to be worse? Answers to these questions depend on how long the agent is expected to play the game; the longer the game lasts, the worse the consequences of prematurely converging on a sub-optimal arm, and the more the agent should explore.

2.1.2 Learning to Predict by the Methods of Temporal Differences

Hereafter we survey the field of reinforcement learning from a computer science perspective [39]. The biggest problem facing a reinforcement learning agent is the *temporal credit assignment*. How do we know whether the action just taken is a good one, when it might have far-reaching effects? One strategy is to wait until the "end" and reward the actions taken if the result was good and punish them if the result was bad. In ongoing tasks, it is difficult to know what the "end" is, and this might require a great deal of memory. Instead, insights from value iteration are used to adjust the estimated value of a state, based on the immediate reward, r, and the estimated value, i, of the next state. This class of algorithms is known as *temporal difference methods* (Sutton, 1988 : [48]) and it is based on an *adaptive heuristic critic algorithm* (Barto, Sutton & Anderson, 1983: [47]).

The adaptive heuristic critic (AHC) algorithm is an adaptive version of policy iteration (Barto, Sutton & Anderson, 1983) in which the value-function is computed by an algorithm called $TD(0)$. A block diagram for this approach is given in Figure 5. It consists of two components: a critic (labeled AHC), and a reinforcement-learning component (labeled RL). The reinforcement-learning component can be an instance of any of the *k-armed bandit* algorithms, modified to deal with multiple states and non-stationary rewards. But instead of acting to maximize instantaneous reward, it will be acting to maximize the heuristic value, v, that is computed by the critic. The critic uses the real, external reinforcement signal to learn to map states to their expected discounted values given that the policy being executed is the one currently instantiated in the RL component. It remains to explain how the critic can learn the value of a policy. We define $\langle s, a, r, s' \rangle$ to be an *experience tuple* summarizing a single transition in the environment. Here s is the environment state before the transition, a is its choice of action, r the instantaneous reward it receives, and s' its resulting state. The value of a policy is learned by using the Sutton's update formula (Sutton, 1988: [48]):

$$\mathbf{V}\left(t+1, s\right) := \mathbf{V}\left(t, s\right) + \alpha\left(r(t) + \gamma\mathbf{V}\left(t, s'\right) - \mathbf{V}\left(t, s\right)\right) \tag{2.1}$$

Namely, whenever a state s is visited, its next estimated value $\mathbf{V}\left(t+1, s\right)$ is updated to be closer to $r(t) + \gamma\mathbf{V}\left(t, s'\right)$, since $r(t)$ is the instantaneous reward received and $\mathbf{V}\left(t, s'\right)$ is the estimated value of the actually occurring next state. The key idea is that $r(t) + \gamma\mathbf{V}\left(t, s'\right)$ is an estimate of the value of $\mathbf{V}\left(t+1, s'\right)$, and it is more likely to be correct because it incorporates the real $r(t)$.

Regarding the convergence of the class of AHC methods, an important result by Williams and Baird [43] is that: under certain conditions, if the learning rate α is adjusted properly (it must be slowly decreased) and the policy (see equation 2.1) is held fixed, then the $TD(0)$ algorithm is guaranteed to converge to the optimal value function.

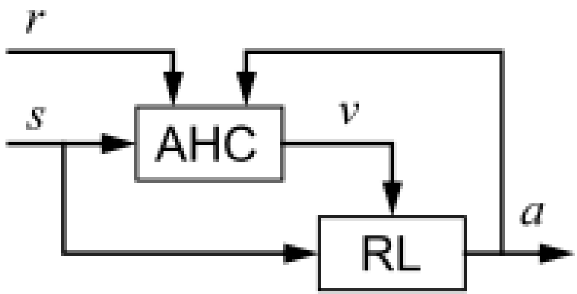

Figure 5: Architecture for the adaptive heuristic critic.

2.2 Game theory: minority game

2.2.1 Reinforcement Learning in a Potential Game

The El Farol bar problem was introduced in 1994 by W. Brian Arthur [18] as a framework to investigate how one models bounded rationality in economics [44]. The original problem was defined as follows:

"N people decide independently each week whether to go to a bar that offers entertainment on a certain night. For correctness, let us set N at 100. Space is limited, and the evening is enjoyable if things are not too crowded - specifically, if fewer than 60 percent of the possible 100 are present. There is no sure way to tell the numbers coming in advance; therefore a person or an agent goes (deems it worth going) if he expects fewer than 60 to show up or stays home if he expects more than 60 to go. Unfortunately, it is necessary for everyone to decide at the same time whether they will go to the bar or not. They cannot wait and see how many others go on a particular Thursday before deciding to go themselves on that Thursday [27]".

Posed in this way, the original problem implies that people are perfectly rational and can therefore make use of deductive reasoning to decide in their actions. But if there was an obvious method that all individuals could use to base their decisions on, then it would be possible to find a deductive solution to the problem. However, no matter what method each individual uses to decide if they will go to the bar or not, if everyone uses the same method it is guaranteed to fail. Therefore, from the point of view of the individual, the problem is ill-defined and no deductive rational solution exists.

In the revisited El Farol bar problem, inductive reasoning was adopted under the assumption that people are bounded rational. Namely, individuals decide whether they will go to the bar or not by employing mental models to predict expected future attendance. In other words they create forecasting models. If an individual using a specific forecasting model predicts attendance to be low then, based on that model, that individual would attend and vice-versa if attendance is predicted to be high. In fact, no forecasting model can be employed by all individuals and be accurate at the same time. Arthur (1994) investigated this model of the El Farol bar problem through the use of computational experiments. The results was that first, mean attendance always converges to the capacity of the bar. Second, on average 40% of the active predictors forecasted attendance to be higher than the capacity level and 60% below. Arthur (1994) expands on these observations by noting that, the predictors self organize into an equilibrium pattern or ecology.

In his work [53], Duncan Whitehead revisited the El Farol bar problem to investigate how one might best model bounded rationality in economics. He began by modeling the El Farol bar problem as a market entry game and describing its Nash equilibria. Then, assuming that agents are boundedly rational in accordance with a reinforcement learning model, he analyzed long-run behavior in the repeated game. The main result he obtained is that, in a single population of individuals playing the El Farol game, learning theory predicts that the population is eventually subdivided into two distinct groups: those who invariably go to the bar and those who almost never do.

2.2.2 The minority game

The minority game [23] is a variant of the El Farol Bar problem and it was proposed by Yi-Cheng Zhang and Damien Challet from the University of Fribourg. This problem embodies some basic market mechanisms, while keeping mathematical complexity to a minimum. The Minority Game is a repeated game where N agents have to decide between two actions, such as buy or sell or attend or not. With N odd this procedure identifies a minority action as that chosen by the minority. Agents who take the minority action are rewarded with one payoff unit. Agents cannot communicate with one another and they have access to publicly available information on the history of past outcomes for a fixed number of periods. As in the El Farol bar problem, from a strategic point of view the problem is ill-defined. Again it is postulated that in such complex strategic interactions [40], agents may prefer to simplify their decision tasks by seeking out behavior rules, or heuristics, that allocate an action for each possible observed history of outcomes. While the El Farol Bar problem was originally formulated to analyze a decision-making method other than deductive rationality, the minority game focuses on the property that no single deterministic strategy may be adopted by all participants in equilibrium. Allowing for mixed strategies in the single-stage minority game produces a unique symmetric Nash equilibrium, which is for each player to choose each action with 50% probability, as well as multiple equilibria that are not symmetric.

Chapter 3

Smart Power Grid Scenario

In this chapter we begin by defining the main issue of our study. Then we focus on the design of power markets.

3.1 Problem definition

A smart grid is an electricity network that can intelligently integrate the actions of all prosumers connected to it, in order to efficiently deliver sustainable, economic and secure electricity supplies. The main issue in this thesis is to propose new solutions which help end-users to optimize their consumption and better manage their own electricity costs. More specifically, the challenge is to make elastic the demand for, and the supply of, electricity of prosumers in order to optimize their energy cost based on power market conditions and on suitable constraints on their power consumption. For this purpose, the efficiency in this work is evaluated according to the ability of users to plan their energy consumption and to minimize their own electricity cost in the long run.

3.2 Architecture of smart power market

According to Robert Wilson's statement [54], the liberalization of infrastructure industries presents classic economic issues about how organization and procedure affect market performance.

In particular, he focuses the analysis on the wholesale power markets. The liberalization of the power industry can be seen as the transition from the monopoly of infrastructures and services to an open and competitive power market potentially of rather different kind. The aim is to replace a tight regulation of vertically integrated monopolies (management of generation, transportation and supply of electricity) with a light regulation of functionally specialized firms, with the supervision of competitive markets. This implies important changes in the management of the power infrastructure and the electricity market [9]. Hereafter by power market 'architecture' we mean a description of the main structural features of a market.

3.2.1 Physical constraints and market model organization

From the viewpoint of standard economic theory [29], wholesale markets for electricity are inherently incomplete and imperfectly competitive [17, 21, 33]. The incompleteness is due to network externalities: the power flow is governed by the Kirchhoff law. In fact the electricity systems are subject to problems of physical coordination of energy flow (inflows and outflows) during dispatching; transmission lines are constrained continuously by operational limits and environmental factors. The competition in the power market is imperfect because the electricity production is capital intensive and construction delays are long compared to variations in supply and demand conditions [30]. On short time scales, electricity prices are inherently volatile and the competition is often imperfect because of technical rigidities on the supply side and the inelasticity of demand.

As consequence of network externalities, an important design issue is the scope of the system operator's authority dedicated to manage the electricity markets. By definition, the system operator is entitled to coordinate, control and monitor the operation of the electrical power system. Given the differences in electricity market structures and regulatory policies around the world, there is no single standard market model. However, any power market can be described in terms of two types of properties:

- **The extent of reliance on market**: every jurisdiction uses a different structure of regulation, governance, system management, and markets. Each market model must handle the basic elements such as the management of the energy market, the transmission and the reserve capacity. Coordination must occur between the electricity market activities and the need to ensure the provision of electricity to end-users [34].

- **The allocation of risks**: financial risk management is often a high priority for participants in deregulated electricity markets due to the substantial price and volume risks that the markets can exhibit. A consequence of the complexity of a wholesale electricity market can be an extremely high price volatility at times of peak demand and supply shortages [22]. The particular characteristics of this price risk are highly dependent on the physical fundamentals of the market such as the mix of types of generation plant and relationship between demand and weather patterns [32, 36].

3.2.2 Integrated market model vs Unbundled market model

The *integrated market model* (see Figure 6 a) is a centralized market where the spot market is coordinated through a system operator commitment. The trading between producers and consumers takes place in a power exchange. The system operator has the power to choose and decide for suppliers the solutions to be used in order to maintain a stable distribution of electricity and their relative costs. By definition, a *pool market* is a market which takes place in a power exchange. It can be subdivided into three subsequent markets: the day-ahead market where the market for energy takes place 24 hours in advance, the intra-day market where the market for energy takes place 1 hours in advance and the real time market where the market for energy takes place 5 minutes in advance.

In the day-ahead and intra-day markets, bids from customers and offers from suppliers are normally firm and they are matched in the market clearing price[1]. The result of the market clearing price becomes an obligation to take and deliver the matched volumes that will be financially settled. Integrated designs start from the premise that, as in traditional power pools, participants are bound together by a relational contract. They employ the system operator as the exclusive manager of all multilateral markets: forward and spot, energy and transmission. The aim is to realize gains from tight coordination in daily operations, and potentially from longer-term obligations and subsidies aimed initially at strengthening overall reliability. But problems may arise because market manipulations by participants have limited counter-measures. As a consequence, integrated designs are most effective when there is vigorous competition, or, if competition is limited, when there is either strong regulation or a legal cartel with ample powers of enforcement. Their advantages are greater when optimization to meet system constraints is more important than participants' flexibility to optimize their own operations and prices on system constraints are more accurate measures of opportunity costs than clearing prices in markets.

In an unbundled market model (see Figure 6 b), the system operator authority (monopoly) is limited to the sale of transmission rights, while trading market strategy, and transmission and reserve capacity management are assigned to each supplier involved in the wholesale market. By definition, a *Forward market* is a sequential market of bilateral contracts of selling energy, buying transmission rights to the transmission system operator and buying energy reserve capacity for real time balancing during the dispatch of electricity. Bilateral contracts is a market mechanism based on physical bilateral contracts. This means that sellers and buyers freely enter into bilateral contracts for power supply. Sellers will normally be generators and buyers will be distribution companies and eligible consumers. This kind of market starts from the oppo-

[1] **Market clearing price**: process aimed at finding the equilibrium price to be paid for the volume of equilibrium determined by the intersection of the aggregate supply curve and aggregate demand curve in a power exchange.

site premise that participation is voluntary, with no long-term relational obligations other than a general tariff approved by the regulator, and that competing forward markets are encouraged to the extent feasible. There is no explicit coordination of the markets for energy, transmission and reserves. Because these markets typically operate in sequence and clear independently, one needs faith in rational expectations to believe they are reasonably efficient. The necessity of a system operator with exclusive authority to manage the public good represented by the transmission system is acknowledged. The system operator's responsibilities include real-time operations that protect system reliability, but its authority to intervene in forward markets is limited to cases where prior commitments promote reliability, such as day-ahead scheduling of transmission. The motives for limiting the scope of the operator's authority are to isolate its monopoly control of transmission from competitive energy markets, and to enable unbundled pricing of energy and transmission. Forward markets can impair efficiency when they are severely incomplete, poorly coordinated, or distorted by regulations. However, ample flexibility and repeated trading opportunities might suffice to approximate markets and to improve coordination.

The two models (integrated market and unbundled market) are equivalent (i.e. they could obtain the same result) without incompleteness of the market model and imperfect competition between producers , consumers and both. Similarly, we can say that the two models (pools and bilateral contracts) are equivalent in a world without transaction costs. In a world with transaction costs however, the bilateral contracts model may result in a sub-optimal outcome, where price and quantity do not reflect real time demand and supply. Instead, in a pool, prices and quantities should reflect actual demand and supply, more so depending on how far ahead of real time the trade occurs. Though prices in a pool may be more volatile than in a contracts market, there are hedging instruments available. In terms of institutional capacity, a simple contracts market is more straightforward and less expensive to set up than a power pool [19].

Hybrid designs enable coordinated markets where forward markets for energy and transmission rights can be unified to capture gains from tighter coordination [24]. The key optimization is a smart market in which prices and resource allocation are obtained from the real trading market model implemented. With these ingredients available, privately organized forward markets have many options for coordinating allocations of energy and transmission rights. A power exchange can conduct a smart market for energy and transmission rights that also includes generators' operating constraints, such as ramp rates and minimum production rates, and auxiliary costs for startup and running, indeed all aspects of unit commitment and scheduling that integrated systems keep within the system operator's control. The hybrid model, from the theoretical point of view, will provide a trade off between two extremes. But it requires on the one hand a priority in the application of one market model at the expense of the other one (precedence for centralized market on forward market or vice versa) and on the other hand the implementation of reliable mechanisms which will prevent the given market organization from market failures. For example, in the New England market pool the centralized market uses a nodal pricing mechanism (price per location) as an incentive for market participants. Nevertheless, the implementation of the smart market requires to cope with the pervasive externalities of the network and with the need of flexibility of the system for risk management.

3.2.3 Power market microstructure

The use of market-based approaches in the electricity system operation continues to mature. In general the electricity market is subdivided into two categories: the wholesale market and the retail sale market. These two markets interact with each other. The wholesale market involves Generator Company (GenCo), Load Service Entities (LSE) or retailer as brokers and distribution companies (DisCo) in a trading market [35]. In the retail market [20], brokers forecast the aggregated resource needs of consumer and shop in the wholesale markets available to them (see Figure 7). Moreover, the wholesale electricity market (also called Spot Elec-

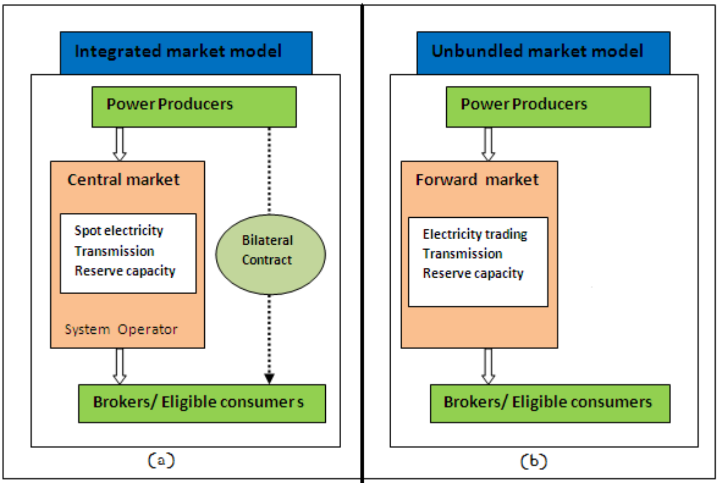

Figure 6: (a) Integrated market model, (b) Unbundled market model

tricity Market) can be subdivided into three subsequent markets:

1. the Day-Ahead Market (DA): which is a market finalized for the exchange of wholesale power between producers, traders and eligible customers. It takes place a day before the effective dispatch of the electricity.

2. the Intra-Day Market (IDM): is an adjustment market on which operators can modify the programs established in response to the Day-Ahead Market submitting further bids on the sale or purchase. It takes place an hour before the effective dispatch of electricity.

3. the Ancillary Services Market (ASM): is the market where the traders make offers of availability of an increase or reduction of the power injected or withdrawn in the grid. This information is used by the transmission system operator to correct programs that violate the limits of transit, and to balance the system in real time against deviations from the program.

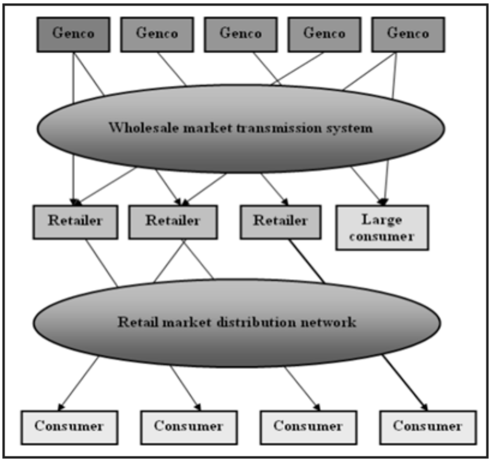

Figure 7: Wholesale and retail sale electricity market

These three markets operate in the pool market, which is under the control of a single operator, who is normally known as Independent Market Operator (IMO). The role of the Independent Market Operator in a wholesale electricity market is to manage the security of the power system in real time and co-ordinate the supply of and demand for electricity, in a manner that avoids fluctuations in frequency or interruptions of supply. The Independent Market Operator service is normally specified in rules or codes established as part of the electricity market. The Independent Market Operator function may be owned by the transmission grid company, or may be fully independent. They are often wholly or partly owned by state or national governments. In many cases they are independent of electricity generation companies (upstream) and electricity distribution companies (downstream). They are financed either by the states or countries or by charging a toll proportional to the energy they carry. The Independent Market Operator is required to maintain a continuous (second-by-second) balance between electricity supply from power stations and demand from consumers, and also ensure the provision of reserves that will allow for sudden contingencies. The Independent Market Operator achieves this by determining the optimal combination of generating stations and reserve providers for each market trading period, instructing generators when and how much electricity to generate, and managing any contingent events that cause the balance between supply and demand to be disrupted. System Operations staff undertake this work using sophisticated energy modeling and communications systems. In addition to its roles of real-time dispatch of generation and managing security, the Independent Market Operator also carries out investigations and planning to ensure that supply can meet demand and system security can be maintained during future trading periods. Examples of planning work may include co-ordinating generator and transmission outages, facilitating commissioning of new generating plant and procuring ancillary services to support power system operation.

Furthermore wholesale transactions (bids and offers) in electricity are typically cleared and settled by the independent market operator (see figure 8). In most of the electricity markets [19], the IMO uses a uniform (or single) clearing price auction in which eligible participants place their bids [10]. The IMO then dispatches the generators from lowest to highest bids until all power demand is met. Each generator that is dispatched is then paid the same price as what was paid to the last unit of electricity needed to meet total demand. The uniform clearing price auction [25] drives generators to reduce their operating costs so that their bids can be lower and, hence, will be accepted. The generators that set the clearing price, and therefore meet the last increment of demand, earn little or no contribution to their fixed costs. The lower cost generators in turn are able to recover some of their long-term debt and other expenses under this auction design. Because the last increment of demand set the clearing price, an explicit price signal to conserve electricity is established. For certain customers who can reduce their demand, a price incentive can be transparently seen.

Hereafter we give an example of how the market clearing price operates in the pool market . The market clearing price is defined by the system marginal price which is the result of the intersection between the aggregated demand and supply curves. The system marginal price is the

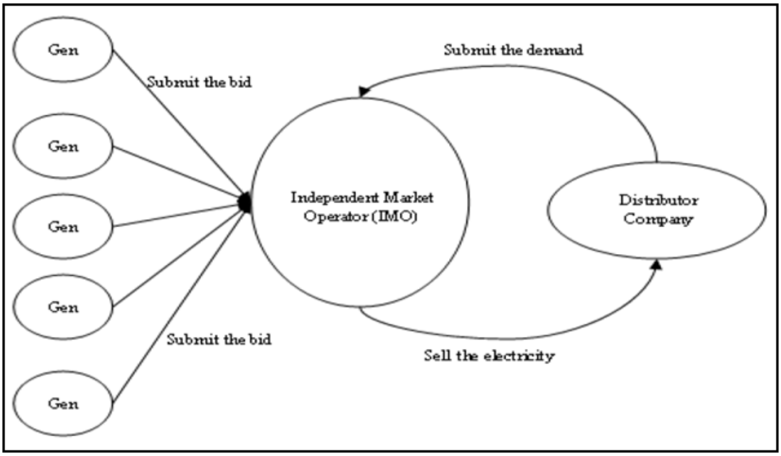

Figure 8: Pool electricity market.

quoted market price paid to all sellers by all buyers; it is the most expensive offer price accepted by buyer to meet their demands. The market clearing price can be subdivided into three steps:

1. Trading participants submit online hourly energy offers (price and quantity): each generating company submits its offer curve; the offer curve is a function assigning exactly one price to each quantity o volume (megawatts) produced. More precisely this price is the marginal cost (MC): the cost of producing one more unit of electricity given the total volume already produced (Q). Hence, each point of this curve (price quantity pair) represents the minimum sale price for the next unit to be produced. Similarly, every buyer submits its bid curve. The bid curve is a decreasing function assigning exactly one price to each quantity (megawatts). Each point of the curve (price quantity pair) represents the maximum purchase price for the next unit to be bought.

2. Matching of the aggregated offer and bid curves: the market operator sums analytically all offer curves (aggregation) and the resulting aggregated curve represents the total amount of energy that

every seller will produce, given the marginal cost (or the selling price) for the next unit of energy. Similarly, all bids curves are aggregated. In this case the curve (see Figure 9) gives the information about the marginal cost (or the purchasing price for the next unit of energy).

3. Price clearing: the intersection point of the two aggregated curves determines the total quantity traded and the equilibrium price. The equilibrium price is often called the winning price. All suppliers who have submitted an offer less than, or equal to, the equilibrium price are allowed to participate in the program production and injection of energy into the grid. they are call in-merit suppliers. Vice-versa, all buyers who have submitted a bid greater than, or equal to, the equilibrium price are allowed to participate in the program for withdrawal of energy from the grid and they are call in-merit buyers (Figure 10).

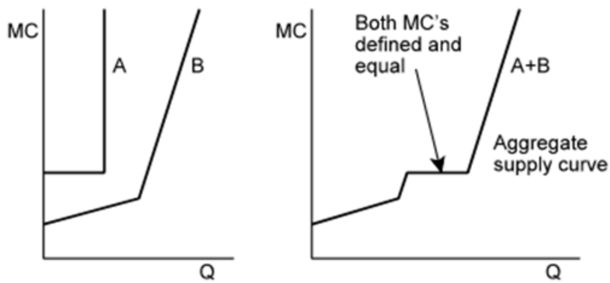

Figure 9: Adding individual supply curves horizontally to find the market supply curve (MC: Marginal Cost, Q:quantity).

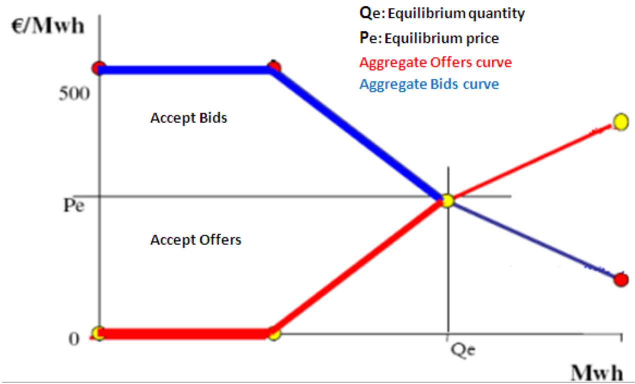

Figure 10: Equilibrium price determination

Chapter 4

DEZENT: a distributed power management and distribution system

In this chapter, we move towards a completely distributed power grid called DEZENT. In what follows, a detailed description of the DEZENT power market is given.

4.1 The idea behind the DEZENT approach

The idea behind the decentralized power management system is to exploit the increasing integration of decentralized energy resources (DER) into the distribution network. DER systems are modern technologies based on solar or wind power, or on other renewables energy sources. In particular we refer to the real-time and distributed power management system DEZENT. DEZENT [49, 50, 51, 52] is the result of a R&D project between the School of Computer Science and the College of Electrical Engineering of Dortmund University, the E.ON Energy company and the German research foundation (DFG). The project was devoted to decentralized and adaptive electric power management through a distributed real-time multi-agent architecture. DEZENT philosophy is to make as local as possible the exchange of electricity between various prosumers. Since the power market is managed in a decentralized manner, prosumers able to carry out more exchanges at the local level will get more benefits.

In the DEZENT R&D initiative, a multi-level bottom-up solution has been developed where autonomous collaborative software agents negotiate available energy quantities (the DEZENT algorithm [49, 50]). Moreover, since much of the production (e.g. from wind farms or from solar panels) is highly unpredictable, the distributed negotiation algorithm of DEZENT is finalized within intervals[1] of 0.5 sec. Such a short interval guarantees that the negotiation situation is considered stable, or constant, for a fair and transparent bidding and offering of partners involved. During each slot of 0.5 sec, the negotiation algorithm proceeds bottom-up, from the 0.4 KV to the 110 KV layers, matching production and consumption of electric power. In addition, a Demand Side and Supply Management (DSSM) algorithm [51] is applied in order to reduce the need of balancing in the power market. At the end of the negotiation, each prosumer independently and simultaneously adjust his/her bidding strategies. This is carried out by the DECOLEARN algorithm [52] which exploits reinforcement learning principles of machine learning.

4.2 Distributed Agent Negotiations in DEZENT

4.2.1 The power grid architecture

The DEZENT power management system focuses on a regional grid where there is a substantial use of renewable energy sources. The power grid architecture (see Figure 11 taken from [50]) is subdivided into four levels. The first level (0.4 KV) is a low-range network covering neighborhoods. The second level (10 KV) is a medium-range area network covering suburbs (regional grid). The third level (110 KV) is a long-distance energy transport network. Finally in the fourth level (380 KV) the electricity

[1] That interval will be called "slot" in our terminology.

is produced from large power plants (coal, gas or nuclear). Most power needs of prosumers are covered through alternative energy sources within the first 2 layers and additional power needs are covered up to the fourth level.

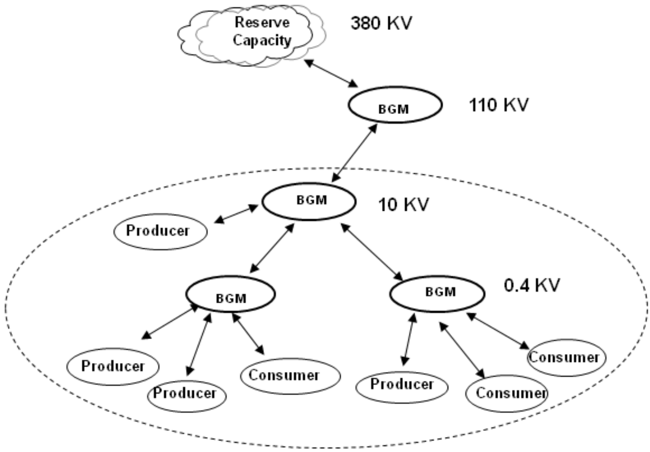

Figure 11: Power grid and associated agents.

At the negotiation layer, the balancing of demand and supply between participants is carried out through Balancing Group Managers (BGMs) which are located in different network layers and operate in parallel on each grid. A BGM is a financial instrument which balances the supply and the demand of electricity between a producer and a consumer who have submitted a similar bid. Moreover, the negotiation will take place in each slot of a day. By definition, a slot is a time interval of 0.5 sec. A day is discretized, resulting in a set of consecutive slots. The distributed negotiations have to be finalized within single intervals. Then the negotiation situation is considered stable, or constant, for a fair and transparent bidding and offering.

During each slot of 0.5 sec, the negotiation algorithm proceeds bottom-up, from the 0.4 KV to the 110 KV layers. The negotiation starts independently for the groups on the lowest level. If a balance cannot be found for all customer agents in a group, then unsatisfied customers are sent to the next-higher BGM and the negotiation scope is extended to that new group of customers. Thus, a slot in DEZENT consist of 3 cycles of negotiations and the sale (purchase) at a fixed cost of the electricity to (from) the main reserve facility. Each cycle consists of 10 rounds of negotiations in which unmatched bids and offers of customer agents are adjusted according to their own negotiation strategies.

4.2.2 Preliminary definition

- *Period*: a period of negotiations has a duration of 0.5 sec. It consists of 3 subsequent negotiations *cycles* followed by the contracting phase at the main reserve layer.

- *Cycle*: a cycle is the negotiation handled by the relevant BGM. It consists of 10 subsequent *rounds* in which the bid of the consumer is forced to increase while the offer of the producer is forced to decrease.

- *Round*: a round is the matching process between fixed bids and fixed offers. During the process, given the current bid of a consumer, *similar* producers are identified for the balancing of the energy needed.

- *Price frame*: $[A_k, B_k]$ is the price frame of the negotiation for the level k ($1 \leq k \leq 3$). Here A_k represents the lower bound of the electricity cost and B_k the upper bound. The bid and the offer of customers are forced to belong to that interval during a cycle of negotiations at level k. The current frame of negotiations at level k, is given by:

$$A_k := A_0 + c(k) \tag{4.1}$$
$$B_k := B_0 - c(k) \tag{4.2}$$

- *Surcharge* $c(k)$: the surcharge is an additional cost paid by each customer at a given level of negotiation. It represents the amortization and maintenance cost of power producers. This cost will increase gradually from the lowest level to the highest one.

$$c(k) := \frac{B_0 - A_0}{2} \cdot Sr \cdot k \tag{4.3}$$

where Sr is a constant, usually of the value of 20%.

- *Bid (cent/Watthours)*: the bid represents the price per unit submitted by a consumer during a round of negotiation n, $n \in [0, 9]$. It is computed according to the bid function

  $$\text{bid}\,(n) = -\frac{1}{e^N} + B_k. \tag{4.4}$$

  where $N = \frac{n}{s_1} + s_2$. A tuple of the form $(s_1, \text{bid}\,(0))$ represents the negotiation strategy set by a consumer agent and the parameter s_2 is determined by the opening bid: $s_2 = -ln\,(B_k - \text{bid}\,(0))$.

- *Offer (cent/Watthours)*: the offer represents the price per unit submitted by a producer during a round of negotiation n, $n \in [0, 9]$. It is computed according to the function

  $$\text{offer}\,(n) = \frac{1}{e^M} + A_k. \tag{4.5}$$

  where $M = \frac{n}{t_1} + t_2$. A tuple of the form $(t_1, \text{offer}\,(0))$ represents the negotiation strategy set by a consumer agent and the parameter t_2 is determined by the opening offer: $t_2 = -ln\,(\text{offer}\,(0) - A_k)$.

- *Allowance (Watthours)*: the allowance is the maximum quantity of energy a consumer is allowed to buy during a negotiation process.

- *Similarity*: a producer is said to be similar to a consumer if the difference between their offer and bid is less than or equal to a constant epsilon ($\text{offer}\,(n) - \text{bid}\,(n) \leq \epsilon$).

- *Negotiated price (cent/Watthours)*: when a consumer C_k and a producer P_k are similar, the negotiated prices per unit of energy is the arithmetic mean of their bid.

  $$\text{price}\,(k) = \frac{\text{bid}_{C_k}\,(n) + \text{offer}_{P_k}\,(n)}{2} \tag{4.6}$$

- *Final unit price (cent/Watthours)*: an additional charge $c(k)$ (see equation 4.3 above) is added to the consumer price and subtracted from the producer fees, respectively. This surcharge guarantees that the most favorable energy prices, for consumers as well as for producers, will be negotiated only on the lowest level.

$$\text{price}_{C_k/P_k} := \frac{\text{bid}_{C_k}(n) + \text{offer}_{P_k}(n)}{2} \pm c(k) \tag{4.7}$$

- *Price for the amount of energy contracted (cent)*: finally the price for the energy contracted is given by :

$$\text{consumerPrice} = \text{amount} * \text{price}_{C_k} \tag{4.8}$$

$$\text{producerPrice} = \text{amount} * \text{price}_{P_k} \tag{4.9}$$

4.2.3 The DEZENT Algorithm

As mentioned in subsection 4.2, a period in DEZENT consists of 3 cycles of negotiation and the sale or purchase of the electricity (at a fixed cost) to the main reserve facility. Each cycle consists of 10 rounds of negotiation in which bid and offer of customer agents are adjusted according to their own negotiation strategies.

Beginning of each round. At the beginning of each round unsatisfied consumers are identified and sorted according to their current bids by the Balancing Group Manager of the level k. Then the consumers are processed top-down starting with the highest bidding consumer. Offers similar to the bid of the first consumer are identified and sorted by price. Offers are processed top-down as well. For closing a contract between the first-listed consumer and the first-listed producer, the needs of the consumer is fulfilled as far as possible within the allowance capacity. This mechanism prevents agent selfishness and avoids, for example, a consumer to purchase a very high amount of energy leaving the other consumers out in the cold. After purchasing a certain amount of watthours from one or more producers, the current consumer's negotiation is interrupted and the algorithm proceeds with the next-listed consumer. After processing the last listed consumer the algorithm starts again with the first interrupted consumer (from the top of the list), allowing it to continue its negotiation. Going through the customer cycle again it proceeds so until no match can be found in the current round any more. Then the algorithm stops and proceeds with the next round. The bids and offers of unsatisfied customers are forwarded to it.

End of each round. At the end of a round we can distinguish the following cases:

1. The needed quantity is only a fraction of the offered quantity. The offer of the producer is adjusted to the difference of the two quantities. The present consumer is deleted. The algorithm proceeds with the next consumer.
2. The two quantities match exactly. Producer and consumer are deleted, and the algorithm proceeds with the next consumer.
3. The needed quantity is not completely covered by the offer. The quantity of the bidder is adjusted to the difference, the producer is deleted, and the algorithm proceeds to identify the next similar producer.
4. If the need of the consumer is not yet satisfied but no similar offers are identified or left, the algorithm proceeds with the next consumer.

End of a cycle. At the end of a cycle (10 rounds of negotiations), unsatisfied customers move up to the next BGM. After that, the negotiation frame is shrunk by a fixed value Sr, typically 20% , thus lowering/raising the upper/lower limits, respectively, by 10% (see Figure 12). An example of the progression of a cycle of negotiation is illustrated in figure 13: there are 6 participating consumers (ascending curves) and 5 producers (descending curves). In the figure, encircled bid/offer pairs (of similar values) refer to contracts and the numbers correspond to the order in which contracts are closed. During the contracting phase, either the consumer curve ends (contract 2) due to needed quantities smaller than those offered, or the producer curve ends (contracts 3, 4) due to offered quantity smaller than those needed. Finally, both curves end when needed and offered quantities match exactly (contracts 1, 5, 6). In this example two consumers remain unsatisfied by the end of the tenth round.

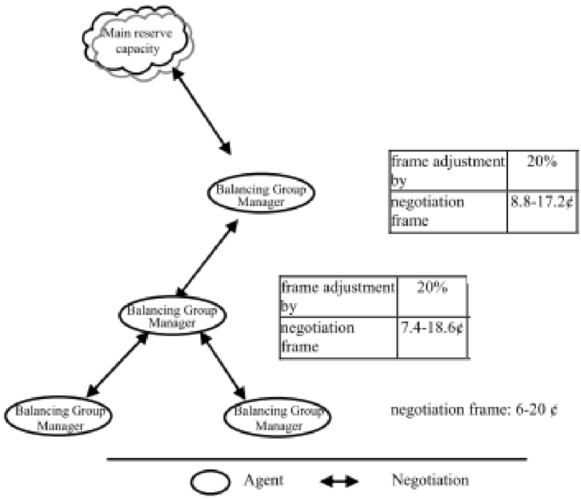

Figure 12: An example of negotiation frame and adjustment.

4.3 Distributed Agent learning in DEZENT

To optimize their behaviors, agents adapt their negotiation strategy at the end of each period. The mechanism used in DEZENT for this purpose is the reinforcement learning principle of machine learning (see Chapter 2). We recall that the negotiation function of a user acting as a consumer (see equation 4.4) is characterized by the tuple $(s_1, \mathrm{bid}\,(0))$ and that of a user acting as a producer (see equation 4.5) is characterized by the tuple $(t_1, \mathrm{offer}\,(0))$. Moreover, the parameter s_1 is chosen

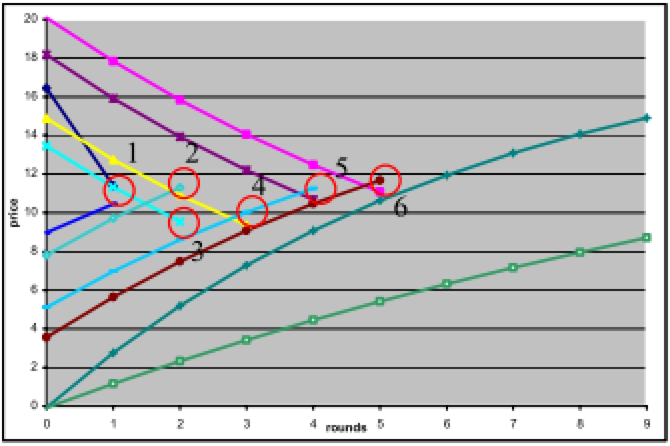

Figure 13: Contracting for energy quantities.

from a finite set of real valued S_C, the parameter t_1 is chosen from a finite set of real valued T_P, the opening consumer bid is chosen from the interval $[A_0, 1/2\,(B_0 + A_0)]$ and finally, the opening producer offer is chosen from the interval $[1/2\,(B_0 + A_0)\,, B_0]$. Let us denote the set of feasible bids for a consumer C by O_C and the set of feasible offers for a producer P by O_P. The strategy space of a prosumer is defined by: $\mathcal{A} := (\mathcal{A}_C, \mathcal{A}_P) := (S_C \times O_C, T_P \times O_P)$. Here the strategy spaces $\mathcal{A}_C$ and $\mathcal{A}_P$ are used when the prosumer acts as a consumer and as a producer respectively.

For the selection of the negotiation strategy of the next slot, 3 modes have been defined in DEZENT: $Exploitation$, $Explore1$ and $Explore2$. $Exploitation$ selects the action with the maximum reward. $Explore1$ randomly picks a strategy which is in the neighborhood of the action with the maximum reward. Finally, $Explore2$ randomly picks any strategy. A mode is randomly determined according to a fixed probability distribution. Then according to the determined mode, a strategy is selected and executed.

At the end of a negotiation slot t, the final achieved price is normalized according to the frame size of the negotiation of DEZENT. More precisely, the unit cost of the electricity $s(t)$ resulting from the negotiation is normalized with respect to the frame size of the negotiation $[A_0, B_0]$.

$$r_C(t) = \frac{B_0 - s(t)}{B_0 - A_0} \text{ for a consumer} \quad (4.10)$$

$$r_P(t) = \frac{s(t) - A_0}{B_0 - A_0} \text{ for a producer} \quad (4.11)$$

Notice that in both cases the parameter varies between 0 and 1, 1 being the best value. Then, the temporal difference method of Sutton [48] is used to derive the reward of the negotiation strategy currently executed. More specifically, let a be a negotiation strategy and $P(t, a)$ be the value of the reward at the beginning of the slot t. Suppose that the strategy a has been executed and let $r(t)$ be the normalized price negotiated under strategy a. The reward $P(t+1, a)$ of the strategy a at the end of slot t is computed by using the Sutton's update formula :

$$P(t+1, a) := P(t, a) + \alpha (r(t) - P(t, a)); \; 0 < \alpha \leq 1$$

Notice that the above formula is simplified with respect to formula (2.1). In fact, the estimate $r(t) + \gamma \mathbf{V}(t, s')$ of the value of the next state s is simplified here as $r(t)$, i.e. $\gamma = 0$, the rational being that the only effect of an action is in this case its reward and not any change of state.

4.4 Demand Side and Supply Management in DEZENT

In DEZENT, the idea of the Demand Side and Supply Management (DSSM) is to further reduce the need of energy balancing in the power grid. The DSSM [51] attempts in a bottom-up fashion at providing the needed reserve energy during undersupply situations; and at channeling any regenerative surplus of energy to those balancing groups where it could be safely stored in real storage (like batteries) or in virtual storage facilities (water heaters, refrigerators). In practice, peak imbalances are smoothened out at the earliest point of time, after each negotiation cycle. At each negotiation level this process is handled in parallel by the involved BGMs. Finally, each BGM reconsiders the situation under the assumption that consumers may give up some portion of their negotiated power quantity, and producers may consider storing some of their excess power.

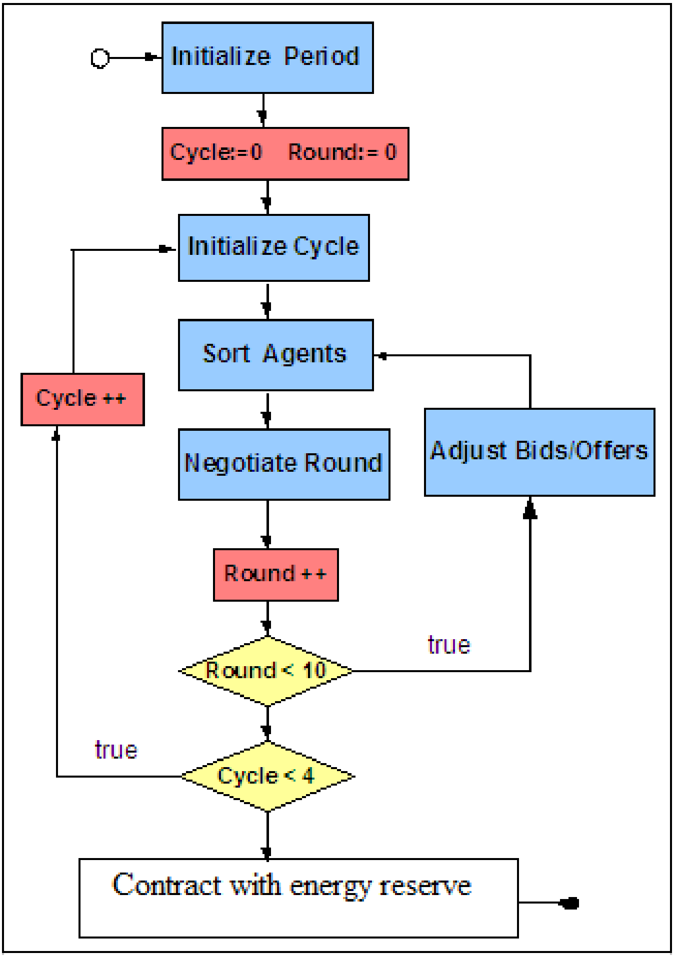

Figure 14: Negotiation period.

Chapter 5

Prosumer Profiling

5.1 Introduction

The aim of our work is to make elastic the demand for, and the supply of, electricity of *prosumers* in order to optimize their energy cost based on power market conditions and on suitable constraints on their power consumption. We believe that independent planning by the prosumers may improve significantly the matching between production and consumption in the DEZENT power grid. In fact, each prosumer should try as much as possible to independently modify his/her power requirements optimizing his cost. In practice, this could mean to help balancing the power market, since the price will favor low consumption/high production when the cost is high and vice versa. Moreover, our approach is not centralized and, in this sense, is different from DSSM. The idea is to exploit the (limited) ability of prosumers of planning in an autonomous way their consumption/production. Hence they do not sign any contract leaving the planning to others: our independent consumer planning is a local matter involving only one prosumer.

More specifically, the prosumer is characterized by the class of consumptions profiles (s)he can adopt during the day. By definition a profile is a function defining consumption in terms of time. We consider a day-long profile (24 hours) and the day is discretized resulting in a set of consecutive slots. The prosumer's objective is to choose, out of the class of consumptions profiles, the profile with the total minimal cost to use for the following day. This choice is carried out by a controller taking into consideration the unit cost of the energy after the DEZENT negotiation. Namely given the class of consumptions profiles and the unit cost of the energy of the previous day, the standard dynamic programming algorithm is used to derive the optimal profile.

In section 5.2 we analyze the power consumption pattern of a power consumer and in section 5.3 we discuss the assumptions on which the controller design is based.

5.2 Power consumption pattern

From the viewpoint of the electricity consumption, captive consumers (households, small businesses, ...etc) cannot actually consume more than a certain amount of available power, even during seasonal drought. Technically, this is called contractually committed power. In the same vein, a normal consumer cannot consume less than a certain threshold because (s)he would like to stay home at a comfortable temperature and to respect the technical constraints of indoor devices. The motivation is that if (s)he consumes more than a certain threshold, then the local distributor may cut off the supply of the electricity. If (s)he consumes less than a certain threshold, then (s)he is losing something in terms of welfare while attempting to save money. However, one of the things (s)he can do is not to give up energy consumption, but to postpone power needs which can be possibly delayed [11, 37, 38, 41, 45]. In this case, the behavior of the consumer is similar to that of a rechargeable battery: the total amount of electricity consumed during an entire day will remain the same, while the amount of electricity consumed during a section of a day might be different.

The main issue is that electricity has a different price depending on the time in which it is used. More specifically, it depends on several exogenous factors, e.g. the actual cost of energy at the global level, but also the existence of convenient energy exchanges at the local level and the fairness of available market mechanisms. The relevant case we consider is that of a decentralized power market like DEZENT where prices can change rapidly and users may not have the explicit information about the cost of the electricity. Hence the behavior (consumption) of the consumer will be based on the information (s)he has on (i) the estimated cost of the electricity and (ii) the consumption constraints. His goal is to minimize the electricity cost at the end of a day while consuming the same amount of energy.

From the viewpoint of a smart house/building (see Figure 15), our solution can be seen as a controller which plays the role of an advanced metering infrastructure. At the physical layer, the controller represents the interface between the power grid and the set of indoor devices of the smart house. At the power market layer the controller will monitor the cost of the energy and will take advantage of this information to plan the energy consumption [31].

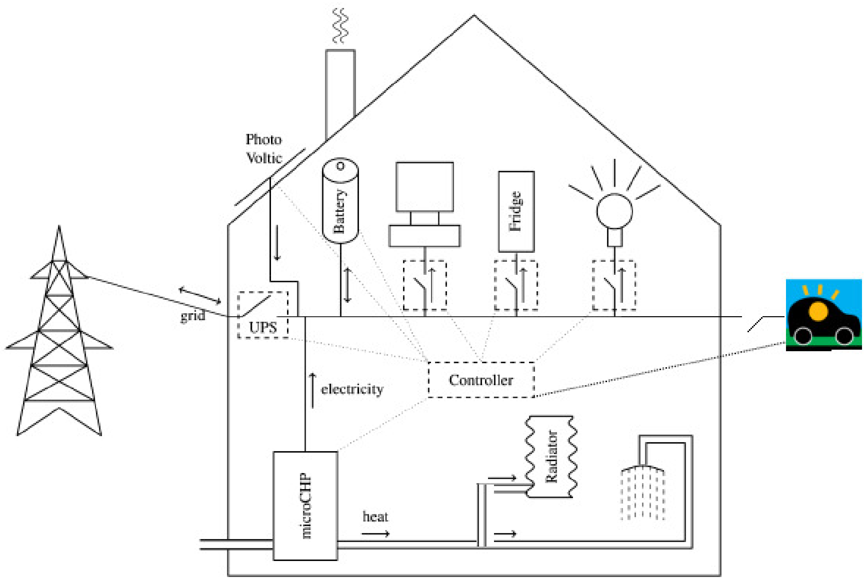

Figure 15: Model of a smart house energy production and consumption optimization.

5.3 Planning phase of the energy consumption

From the point of view of the prosumer, the unit cost of energy after the DEZENT negotiation depends mainly on two factors: (a) the free market power cost (energy reserve); (b) the prosumer environment: heavy production, heavy consumption and equilibrated prosumer population may yield rather different energy costs. The point (a) accounts for the fact that the electricity may have a rather different price depending on the time period, be it of the day, of the week or of the year. The variance of the electricity cost in different periods of the day is particularly relevant: high variance does offer the prosumer a better chance of allocating his/her power needs when it is more convenient economically. On point (b), we observe that prosumers who are not satisfied in their sub layers are lifted up on the next layer and this might result in a less profitable cost of the electricity. From now on, we will consider only the case of users which are consumers.

The case of producers is for most aspect dual. The difference between consumer and producer can be represented in our approach by different constraints, in particular by the size of the energy storage media. The producer with lots of intermittent sources will have a limited flexibility. However, energy sources like cogeneration systems (CHP) for the production of heat and power and biogas could still provide consistent energy reserves.

A consumer is characterized by the class of consumption profiles (s)he can adopt during the day. Also, it is important when (s)he has to choose a particular profile: (i) at the beginning of the day or (ii) slot by slot. Furthermore, in the former case it must be decided what happens if the actual consumption in a slot turns out to be different than the planned consumption: the profile for the rest of the day should be replanned or not? In our model, we characterize the allowed profiles as variations with respect to the ordinary consumptions. They must satisfy the following constraints:

1. the energy variation in a slot has a lower and an upper bound;

2. the energy consumed in the whole day is fixed, i.e. if in some slot the consumption is lower than average, in some other slot it must be higher. This constraint allows to delay (or to anticipate) a job (e.g. a laundry washing cycle) but not to abolish it; thus the sum of all variations must be 0.

3. summing up all the variations from the beginning of the day to any time, we cannot exceed a lower and an upper bound. This constraint accounts for available energy storage media, like electric vehicle batteries or thermic accumulations due to anticipated heating, or delayed air conditioning.

Given the class of consumption profiles, the consumer will choose the optimal profile on the basis of the information it has on the unit cost of the energy. Of course he/she cannot foresee the future and will rely on the information of past days. If the prosumer knew the actual cost of energy in every slot, he/she could compute (by dynamic programming) a profile of variations which, among the allowed profiles, would optimize the gain. A reasonable estimate of the cost of energy in every slot can be obtained by looking at the values in the previous day. In fact, if we assume that the free market cost in the same slot of the previous day is the same, that the prosumer environment is the same and that the DECOLEARN algorithm is close to convergence, we can safely rely on the after-DECOLEARN costs of the previous day. Of course the prosumer environment has a stochastic behavior, thus it might behave differently in the two days. To improve the estimate, we could take an average of several days, or we could install a new reinforcement learning process. However both approaches apparently do not offer significant improvements experimentally, while in addition the latter choice is computationally very expensive, since different slots should be equipped with different parameters. In short, starting from a certain profile of the power cost in the free market, which works as a scale factor, and from a certain prosumer environment (heavy production, heavy consumption or equilibrated population), the DEZENT simulator is used to extract the power cost at the prosumer level in every slot. Then dynamic programming is used in a standard way to optimize cost reduction within the allowed constraints. Finally the resulting profile is applied the following day. Moreover as we assume that the optimization problem is largely independent from the DEZENT negotiation, the planning phase and the negotiation phase could operate with different timings, the former every day and the latter every slot.

In summary, our approach depends on the following parameters: (a) the profile of the free market power cost in every slot, and in particular its variance; (b) the prosumer environment: heavy production, heavy consumption or equilibrated prosumer population; (c) how to estimate the cost of energy after the DEZENT negotiation; (d) the amount of allowed power increase/decrease; and (e) the available reserve. Items (a) and (b) affect the simulations but their values are not directly available to the prosumer, while items (c), (d) and (e) are available inputs for prosumers optimal planning.

Furthermore as at any moment the consumers can modify their plans for whatever reason we study the problem of (sub) optimal repeated replanning for the rest of the day when consumption varies with respect to the anticipated values. These changes can be due to the need of additional power - which however does not require any changes of the planned profile - or to the loss of a part of the available energy reserve.

5.4 Formal description of the problem and of the proposed algorithm

Hereafter, the optimization problem and the proposed dynamic programming algorithm used to solve it are defined.

Notations: some definitions and notations are listed below ($\mathbf{N}$ are the natural numbers).

Discretized energy	:	$s_1, \ldots, s_n$ slots in a day, $n : \mathbf{N}$
	:	$e : \mathbf{R}$ basic energy level, $e > 0$
	:	ae average consumption, $a : \mathbf{N}$
	:	re maximal energy reserve, $r : \mathbf{N}$
	:	$r_0 e$ initial energy reserve, $r_0 : \mathbf{N}$
	:	$\pm ke$ maximal variation in energy consumption, $k : \mathbf{N},\ k \leq a/2$
	:	$o : \mathbf{Z} \to \mathbf{R}$ overhead, if $0 \leq x$ then $o(x) = x$ else $x \leq o(x) \leq 0$
Unitary energy cost	:	$c_i : \mathbf{R}$ in slot $s_i, \quad i = 1, 2, \ldots, n \quad c_i \geq 0$

Optimization problem: the optimization problem is then defined by the decision variables, the function to be minimized and the constraints on the energy consumption.

Decision variables	:	$-k \leq x_i \leq +k,\ x_i : \mathbf{N}$ where x_i is the variation for slot $s_i,\ i = 1, 2, \ldots, n$
Cost function to be minimized	:	$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} (o(x_i) + a)c_i$
Optimal cost	:	$C = \min_{x_1,\ldots,x_n} \sum_{i=1}^{n} (o(x_i) + a)c_i$
Constraints	:	$\forall\, 0 \leq j \leq n.\ \ 0 \leq r_0 + \sum_{i=1}^{j} x_i \leq r$
	:	$\sum_{i=1}^{n} x_i = 0$

Algorithm: the proposed solution algorithm decomposes the problem into subproblems, so that an efficient dynamic programming approach can be employed. Let $C_j(y_j) : \mathbf{R} \cup \{\infty\},\ j = 0, \ldots, n,\ 0 \leq y_j \leq r$ be the optimal energy costs for slots $s_1, \ldots, s_j$, when the final energy reserve at slot s_j is $y_j e$. Here $C_j(y_j) = \infty$ if energy reserve $y_j e$ cannot be achieved at slot j. Thus $C_0(y_0)$ (no slot has elapsed yet) is everywhere ∞ except for $C_0(r_0) = 0$.

$$
\begin{array}{lll}
\text{Subproblems} & : & C_j(y_j) = \min\limits_{x_1,\ldots,x_j} \sum\limits_{i=1}^{j} (o(x_i) + a)c_i \quad j = 1, 2, \ldots, n \\
 & : & \forall i'.\, 1 \leq i' \leq j, \qquad 0 \leq r_0 + \sum\limits_{i=1}^{i'} x_i \leq r \\
 & : & r_0 + \sum\limits_{i=1}^{j} x_i = y_j \qquad 0 \leq y_j \leq r
\end{array}
$$

$$
\begin{array}{lll}
\text{Dynamic} & & \\
\text{programming} & : & C_j(y_j) = \min\limits_{\substack{-k \leq x_j \leq k \\ 0 \leq y_j - x_j \leq r}} C_{j-1}(y_j - x_j) + (o(x_j) + a)c_j, \\
 & & j = 1, 2, \ldots, n \\
 & : & C_0(y_0) = \text{ if } y_0 = r_0 \text{ then } 0 \text{ else } \infty \\
 & : & C_n(r_0) = C
\end{array}
$$

The value of C_j at slot j can be computed sequentially in terms of C_{j-1} by looking backwards for $C_j(y_j)$ to the optimal energy costs at slot $j - 1$ for eligible values $y_j - x_j$ of the energy reserve.

Finally, an *optimal strategy* S is any sequence $S = (\widehat{x_1}, \widehat{y_1}), \ldots, (\widehat{x_n}, r_0)$ such that the values of $\widehat{x_j}$ and of $\widehat{y_{j-1}}$ are computed backwards from $\widehat{y_j}$, $j = n \ldots, 1$, by letting $\widehat{y_n} = r_0$, the final reserve being r_0. Formally:

$$
\begin{array}{lll}
\text{Optimal strategies} & : & C_j(\widehat{y_j}) = C_{j-1}(\widehat{y_j} - \widehat{x_j}) + (o(\widehat{x_j}) + a)c_j,\ j = 1, 2, \ldots, n \\
 & : & \widehat{y_n} = r_0
\end{array}
$$

The time and space complexity of the algorithm are $O(nrk)$ and $O(nr)$ respectively.

Chapter 6

Experimental results

6.1 Controller Algorithm

The control algorithm has two inputs: (i) the definition of the class of allowed consumer profiles; and (ii) the cost of a unit of energy which resulted by the DEZENT negotiation in each slot of the previous day. The output of the algorithm is an allowed profile assigning to each slot an energy variation of minimal total cost. The complexity of such an algorithm depends on the definition of the class of allowed profiles. For the consumer characterization mentioned above in section 5.3, an efficient dynamic programming algorithm can be defined: the subproblems are the subprofiles (from the beginning of the day to a given slot) of minimal cost for all the sums within the capacity of the energy reserve . The complexity of the algorithm is quite good: it is linear in the number of slots in a day, in the number of allowed sums and in the number of levels (here five) of variation with respect to the ordinary energy consumption. In what follows we present the operation of the controller algorithm.

A Java programming language has been used for the implementation of the optimal controller and the available DEZENT simulator[1] has been extended to it.

[1]The DEZENT simulator has been kindly made available by the DEZENT consortium, and deployed on the IMT cluster.

6.2 Consumer characterization

The set of allowed consumption profiles could be defined in more flexible ways, accounting for different classes of consumers and different parameters could be chosen for the consumer characterization. In our case, for every slot the allowed power variation with respect to the ordinary consumption has five levels: -2Δ, $-\Delta$, 0, $+\Delta$ and $+2\Delta$. The value of the deviation Δ is 25% of the average energy consumption. If the consumer has a bounded energy reserve, the available reserve capacity is between -10Δ and 10Δ. Moreover, each time the consumer uses the energy reserve, an additional cost is added [12]: 10 % more for the action $-\Delta$ and 15% more for -2Δ. This constraint accounts for the fact that not all the energy stored can be made available and a part of it is lost by the Joule effect or by other kinds of energy transformation losses.

6.3 Space of the experiments

We performed extended experimental studies based on the available DEZENT simulator and on the Java implementation of the optimal controller. An experiment depends essentially on three parameters:

1. the free market power cost, which can exhibit high, low or average variance: for this we chose real data from the day ahead market prices of Germany [13, 14], Italy [15] and Switzerland [16] respectively (see Figures 16, 17 and 18);

2. the prosumer environment, namely heavy production, heavy consumption or equilibrated prosumer population, in which the total amount of the electricity produced in the subnet is respectively greater than, less than and equal to, the total amount needed in the subnet. In the heavy consumption situations, the additional, needed power is made available at the large power plant level, at a price which depends on the time of the day. Analogously for the equilibrated population and heavy production situations. In

all these cases, the profile cost of the electricity at the global level (namely at the large power plant level) was the same for all days.

3. the available energy reserve capacity which is considered either finite or infinite.

For every combination of the parameters we synthesized the optimal controller and computed the gain with respect to a neutral consumer operating in the same environment, i.e. within a context of other prosumers (not only consumers), which are the same in both cases. Furthermore we expected that, the highest energy cost reduction is obtained when we have a high variance on the profile cost of the electricity, the prosumer environment is in the heavy consumption situation and the energy reserve capacity of the controller is infinite. Vice versa, the lowest energy cost reduction corresponds to low variance, oversupply situation and finite reserve capacity.

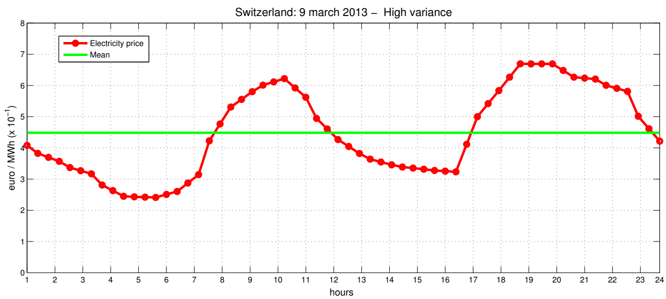

Figure 16: Day ahead power market: Switzerland, 9 march 2013 - High variance.

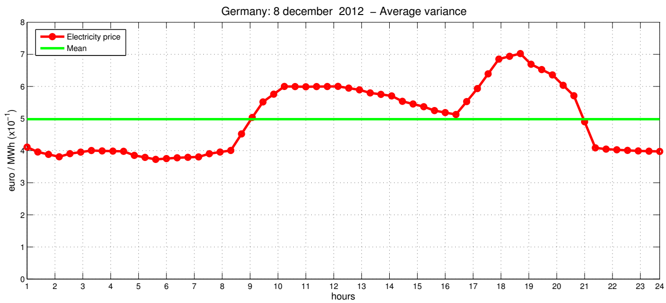

Figure 17: Day ahead power market: Germany, 8 december 2012 - Average variance.

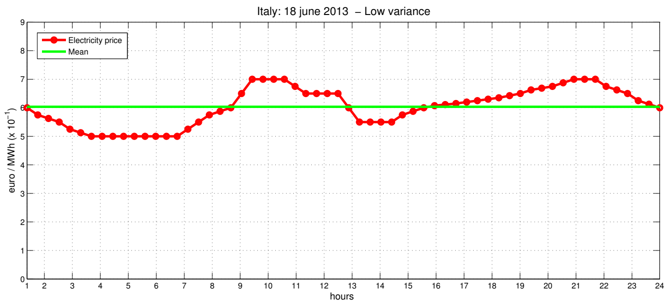

Figure 18: Day ahead power market: Italy, 18 june 2013 - Low variance.

6.4 Experimental studies

The experiments were conducted on the IMT cluster at the IMT Institute of Advanced Studies Lucca, simulating a 3 day service period (see Table 1) and our comparative studies were based on the total cost of the electricity paid at the end of the last day. The performance of the controller relies on how stable the negotiation process is in different days, which in turn depends on the stochastic evolution of the other involved partners. Here the last day has been considered, since in this way transitory effects are minimized. In fact during the first day simulation corresponding to 60 steps, the reinforcement learning component of the DEZENT

Table 1: Experimental Setup of the placebo test

Architecture	Negotiation Levels	1
	BGM on Level 1	1
	Clients	15
	Producers ($50-200$ KW)	10
	Consumers (200 KW)	5
Electricity price	Day duration: 60 slots	
	Profile cost of the electricity (free market)	
Prosumer environment	Heavy consumption	
	Equilibrated prosumer population	
	Heavy production	
Energy reserve	Infinite	
	Finite: -10 to 10	
Controller	Class of consumption profile	
	Planning phase: optimization	
Simulations	Duration: 3 days	
	Test 1: active consumer	
	Test 2: neutral consumer	

simulator will gather as much information as possible and will become stationary.

The first group of simulations concerns the high variance of the free market power cost (day ahead power market: Switzerland, 9 march 2013) along with the infinite reserve capacity of the optimizing controller. Figures 19 and 20 represent the result of the simulation in the heavy consumption case of the prosumer population. Figure 19 synthesizes the operation of the controller during the day 2 and 3 and Figure 20 compares the electricity cost achieved at the end of the day 3. The two upper curves of figure 19 represent the unitary cost of energy as resulting from the negotiation phase at day 2 (curve in black color) and at day 3 (dashed curve in blue color). The difference between the two upper curves gives an idea of the possible variations between the outcomes of different negotiations. Notice that the profile of the global energy cost and the context of competing prosumers is the same in both days. The two lowest curves of figure 19 represent the result of the optimization algorithm applied to the curve of day 2 (curve in red color) and to the curve of day 3 (curve in green color). The curves plot the sum (from the beginning of the day) of the suggested variations with respect to the ordinary energy consumption: according to the constraints we assumed on the consumption profiles, the sum must stay between the available energy reserve capacity and should end up at 0, i.e. the day long energy consumption should be left invariant. We can notice that the controller correctly suggests variations which are opposite with respect to the negotiated cost. The three curves in Figure 20 report the actual energy cost for day 3 of a neutral consumer (upper, red curve), of a consumer employing the optimizing controller (lower, blue curve) using the negotiated cost of day 2, and of a consumer able to foresee the future (lower, green curve). In the latter case the optimization algorithm is applied to the negotiated cost of the same day 3. However, the two negotiated cost profiles are similar enough to guarantee a remarkable economic advantage for the active consumer. In fact the curves of the active consumer (lower, green and blue curve) are close together while they are far away from that of the neutral consumer (upper, red curve). Similarly Figures 21 and 22 show the results of the negotiation in the equilibrated prosumer population case while Figures 23 and 24 concern the heavy production case. In general we see the same informations as in the heavy consumption case mentioned above. Furthermore if we compare the gain realized, in terms of cost reduction, between the three prosumer population cases (heavy consumption, equilibrated population and heavy production) then we can observe that the gain realized decreases as we switch from the heavy consumption case to the equilibrated population case, and from the equilibrated population case to the heavy production case. This result accounts for the fact that the power cost in the free market is higher than, or equal to, any possible negotiated price in the subnet in DEZENT. Hence the positive effect of the optimization procedure is more remarkable when the prosumer context is heavy consumption. In fact in this case the effect of the cost differences in the free market power are not obscured by the effect of the negotiations.

The second group of simulations is similar to the first one (day ahead power market: Switzerland, 9 march 2013 - high variance) except for the fact that the energy reserve capacity of the optimizing controller is finite. The pairs of figures (25, 26), (27, 28) and (29, 30) are the result of the negotiation in the heavy consumption, equilibrated population and heavy production case respectively. As we can see, the active consumer again has spent less than the neural consumer in each case of the prosumer environment. Nevertheless if we compare the electricity cost achieved at the end of the day 3 between the active consumer with an infinite energy reserve capacity and that of the active consumer with a finite one, then we can see that the active consumer with a finite energy reserve capacity has spent more than the other. Namely the capacity of the energy reserve influences the performance of the optimizing the controller: the higher the capacity, the better is the allocation of the resource.

The third (see Figures from 31 to 36) and fourth (see figures from 37 to 42) groups of simulations concern the case of the average variance of the electricity cost (day ahead power market: Germany, 8 december 2012). Finally the fifth (see Figures from 43 to 48) and sixth (see Figures from 49 to 54) group of simulation corresponds to the case of low variance of the electricity cost (day ahead power market: Italy, 18 june 2013). Three observations can be made as result of the comparison between the active and the neutral consumer:

1. The energy cost reduction increases as we switch from the low variance to the high variance of the free market power cost.

2. The energy cost reduction increases as we switch from the heavy consumption situation to the heavy production situation of the prosumers environment.

3. The energy cost reduction increases as the energy reserve capacity of the optimizing controller switch from the finite case to the infinite one.

In summary, the combination of those three results leads to the definition of the best and worst case scenario. Namely the highest energy cost reduction was obtained when we have a high variance on the profile cost of the electricity, the prosumer environment is in the undersupply situation and the reserve capacity of the controller is infinite (see Figure 20). Vice versa, the lowest energy cost reduction corresponds to low variance, oversupply situation and finite reserve capacity (see Figure 54).

Study of a mixed prosumer population.
In these simulations, the population of agents consists of 10 power producers and 5 power consumers, where 2 consumers are always neutral, 2 are always active and the last one corresponds to the placebo agent test which changes the role from being neutral in the first simulation and active in the other. The free market power cost is that of Switzerland which exhibits high variance. The available energy reserve capacity of an active consumer is considered infinite. Figures 57 and 58 represent the result of the simulation in the heavy consumption case of the mixed prosumer population. Similarly Figures 59 and 60 show the results of the negotiation in the equilibrated prosumer population case while Figures 61 and 62 concern the heavy production case. The significance of these simulations will be discussed in the conclusion.

(Sub) optimal repeated re-planning: loss of the energy reserve.
In order to study the impact of the loss of the energy reserve on the performance of the optimizing controller, we have considered the following case: Germany (december 8, 2012), heavy consumption situation of the prosumer environment and infinite energy reserve capacity of the controller. We have simulated 2 types of energy loss in the same prosumer environment: one happening at 9.75 o'clock (negotiation number 23) and the other one at 12.5 o'clock (negotiation number 30). Hence the energy reserve was not anymore available during the negotiation number 23, 30 respectively; and the controller has optimally replanned the energy consumption for the rest of the day. Namely the sum of all the deviations starting from the negotiation number 23, 30 respectively, to the end of the day will be equals to 0. Then we have compared the result of their energy costs achieved at the end of day 3 with respect to an active consumer who is not subject to any energy loss.

Figure 55 concerns the profiles of the energy used during the day 3. The profiles are the result of the (sub)optimal re-planning that has happened at 9.75 o'clock (black curve) and 12.5 o'clock (blue curve) respectively. We can see that the sum up of the deviation from the beginning of the day does not end up to zero in both cases. The small black circle indicates the level of the energy consumption just before the re-planning phase took place at 9.75 o'clock. In fact at the end of the day, the value of the energy reserve end up again to the same level (of value 38). Similarly, the small blue circle indicates the level (of value 24) of the energy consumption just before the re-planning phase took place at 12.5 o'clock.

In Figure 56, we have plotted the energy cost for day 3 of : (i) the active consumer (upper, black curve) with energy loss happening at 9.75 o'clock, (ii) the active consumer (middle, blue curve) with energy loss happening at 12.5 o'clock and (iii) an active consumer (lower, red curve) without energy loss. We can see that the highest energy cost has been paid by the the consumer who has lost more energy (upper, black curve) while the lowest one (lower, red curve) corresponds to the consumer who was not subject to any loss of the energy reserve. In summary the consumer who has lost more energy has spent more at the end of the day. Hence during the re-planning, the additional cost paid is proportional to the amount of energy lost.

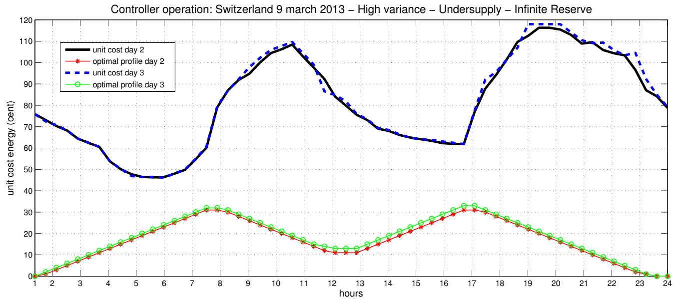

Figure 19: Switzerland, 9 march 2013 - high variance - undersupply - infinite reserve: planning phase.

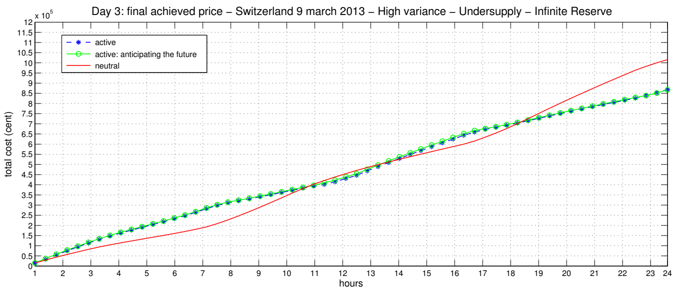

Figure 20: Switzerland, 9 march 2013 - high variance - undersupply - infinite reserve: active vs neutral.

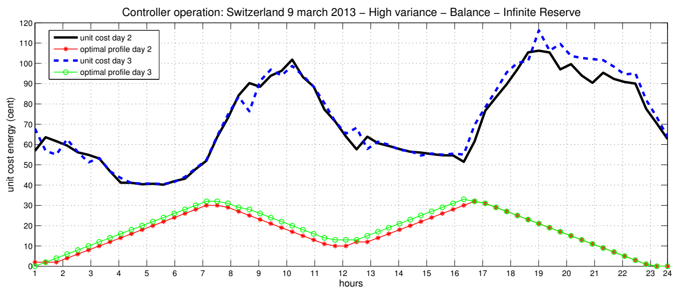

Figure 21: Switzerland, 9 march 2013 - high variance - balance - infinite reserve : planning phase.

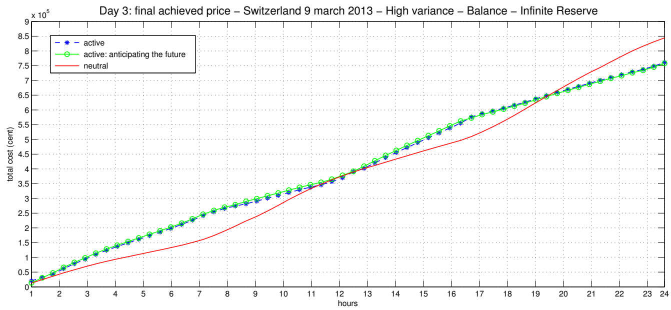

Figure 22: Switzerland, 9 march 2013 - high variance - balance - infinite reserve: active vs neutral.

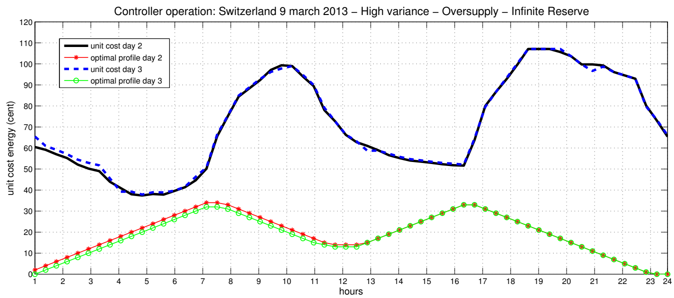

Figure 23: Switzerland, 9 march 2013 - high variance - oversupply - infinite reserve: planning phase.

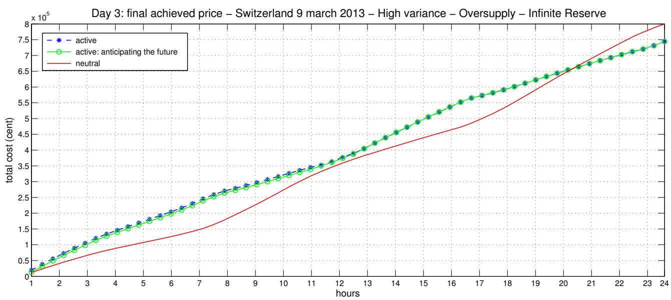

Figure 24: Switzerland, 9 march 2013 - high variance - oversupply - infinite reserve: active vs neutral.

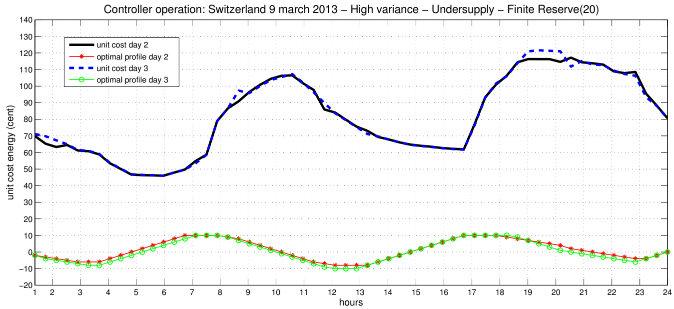

Figure 25: Switzerland, 9 march 2013 - high variance - undersupply - finite reserve: planning phase.

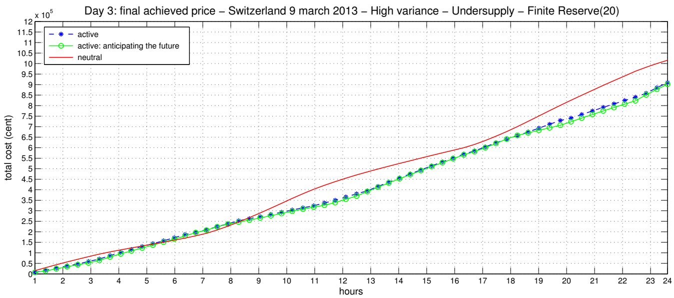

Figure 26: Switzerland, 9 march 2013 - high variance - undersupply - finite reserve: active vs neutral.

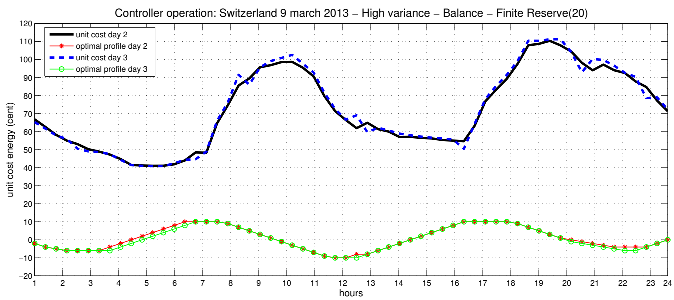

Figure 27: Switzerland, 9 march 2013 - high variance - balance - finite reserve : planning phase.

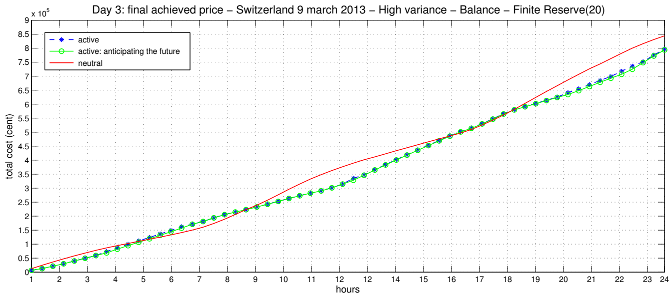

Figure 28: Switzerland, 9 march 2013 - high variance - balance - finite reserve: active vs neutral.

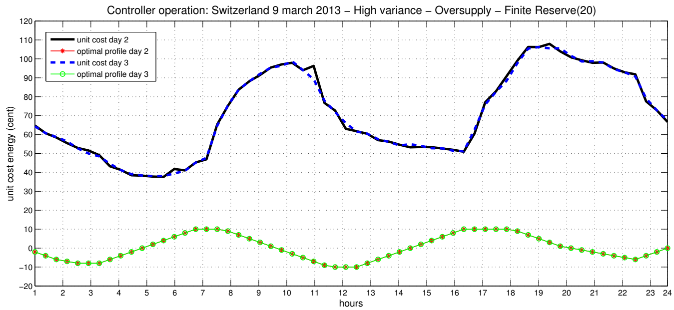

Figure 29: Switzerland, 9 march 2013 - high variance - oversupply - finite reserve: planning phase.

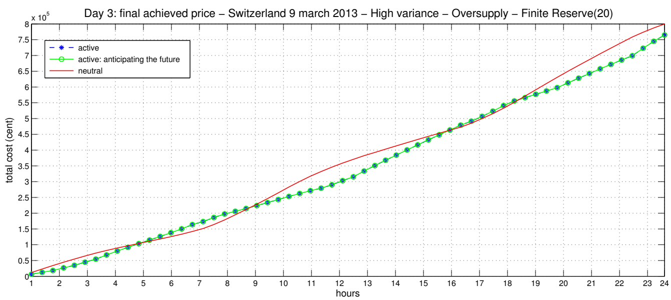

Figure 30: Switzerland, 9 march 2013 - high variance - oversupply - finite reserve: active vs neutral.

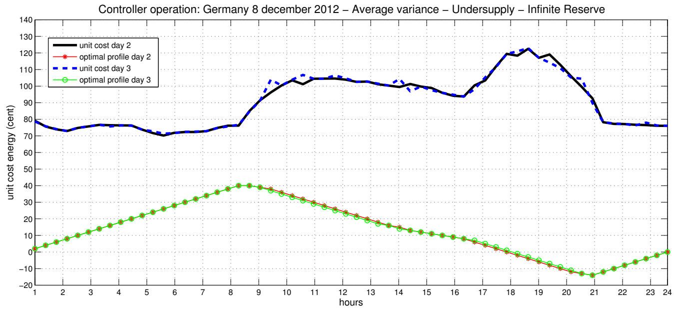

Figure 31: Germany, 8 december 2012 - average variance - undersupply - infinite reserve: planning phase.

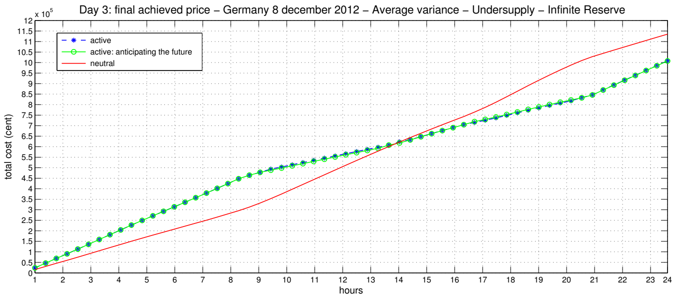

Figure 32: Germany, 8 december 2012 - average variance - undersupply - infinite reserve: active vs neutral.

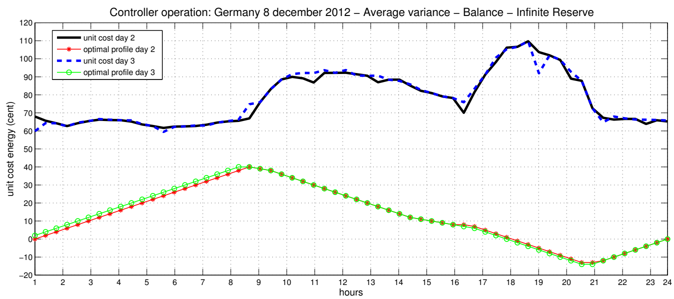

Figure 33: Germany, 8 december 2012 - average variance - balance - infinite reserve: planning phase.

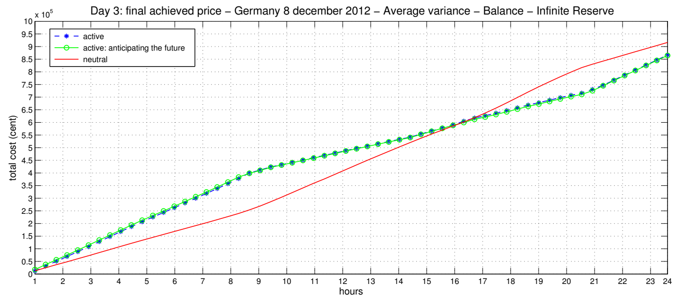

Figure 34: Germany, 8 december 2012 - average variance - balance - infinite reserve: active vs neutral.

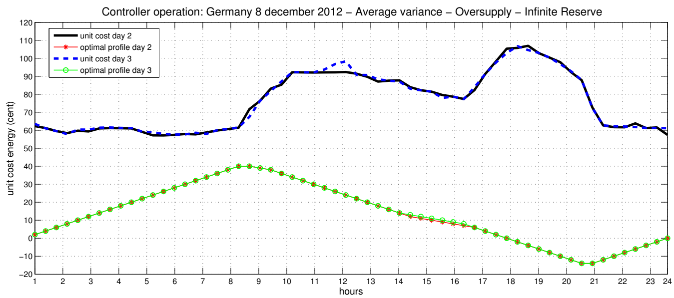

Figure 35: Germany, 8 december 2012 - average variance - oversupply - infinite reserve: planning phase.

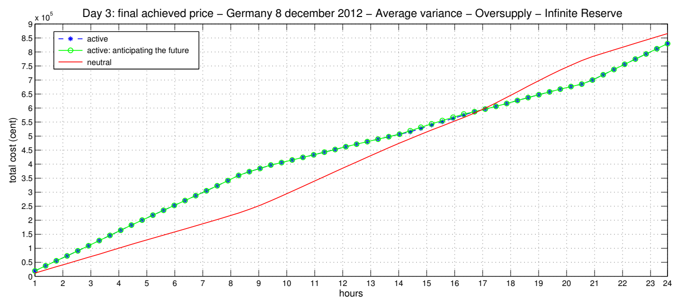

Figure 36: Germany, 8 december 2012 - average variance - oversupply - infinite reserve: active vs neutral.

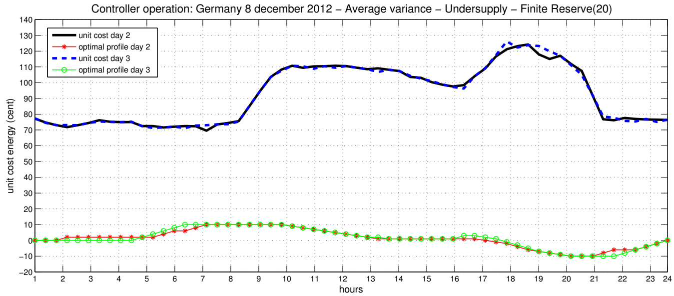

Figure 37: Germany, 8 december 2012 - average variance - undersupply - finite reserve: planning phase.

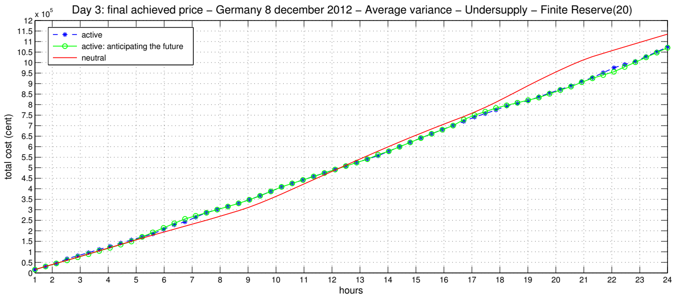

Figure 38: Germany, 8 december 2012 - average variance - undersupply - finite reserve: active vs neutral.

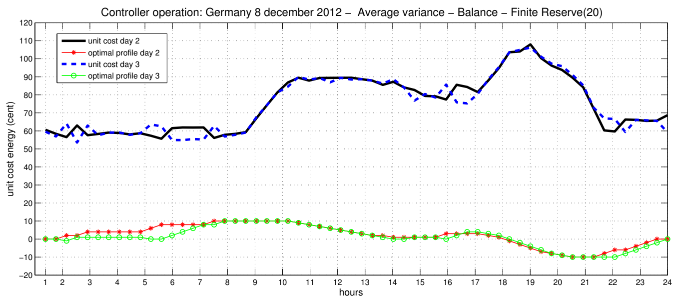

Figure 39: Germany, 8 december 2012 - average variance - balance - finite reserve: planning phase.

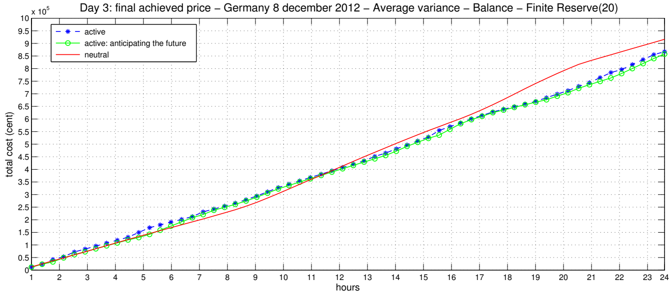

Figure 40: Germany, 8 december 2012 - average variance - balance - finite reserve: active vs neutral.

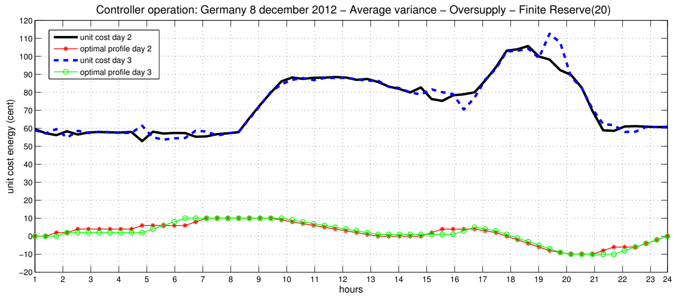

Figure 41: Germany, 8 december 2012 - average variance - oversupply - finite reserve: planning phase.

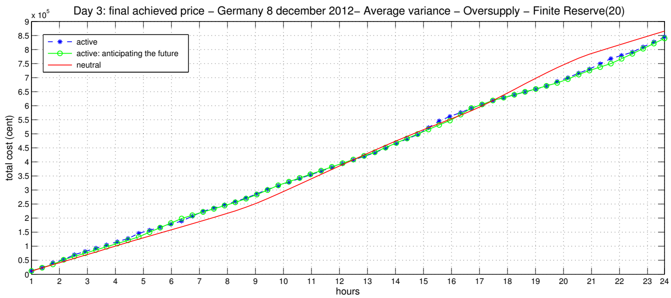

Figure 42: Germany, 8 december 2012 - average variance - oversupply - finite reserve: active vs neutral.

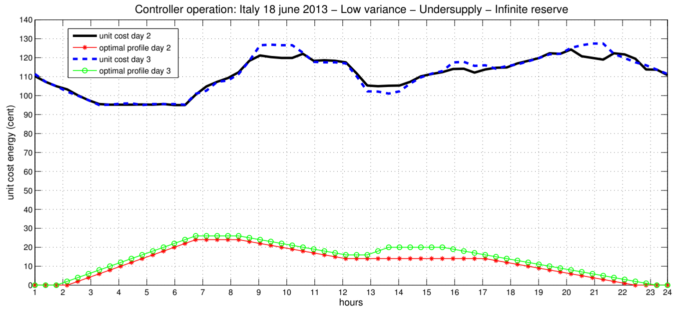

Figure 43: Italy, 18 june 2013 - low variance - undersupply - infinite reserve: planning phase.

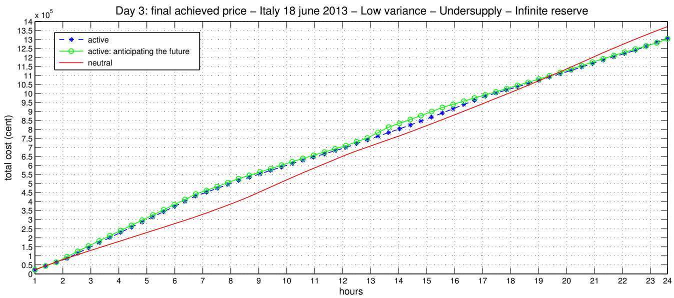

Figure 44: Italy, 18 june 2013 - low variance - undersupply - infinite reserve: active vs neutral.

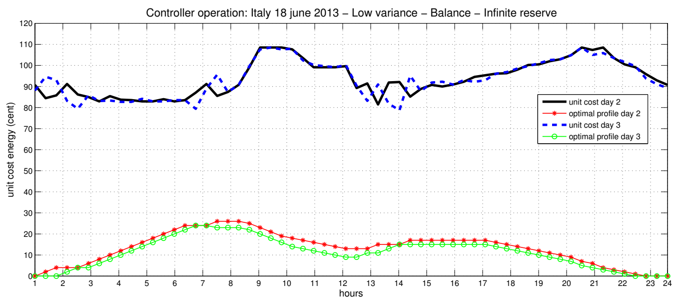

Figure 45: Italy, 18 june 2013 - low variance - balance - infinite reserve: planning phase.

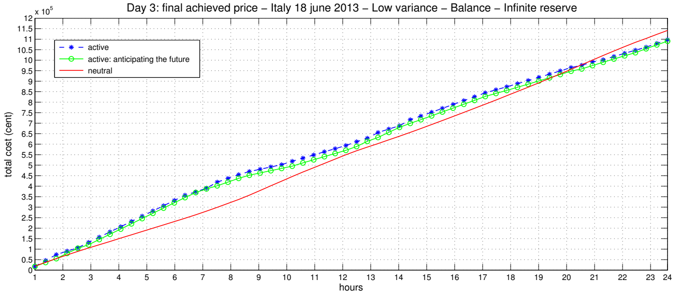

Figure 46: Italy, 18 june 2013 - low variance - balance - infinite reserve: active vs neutral.

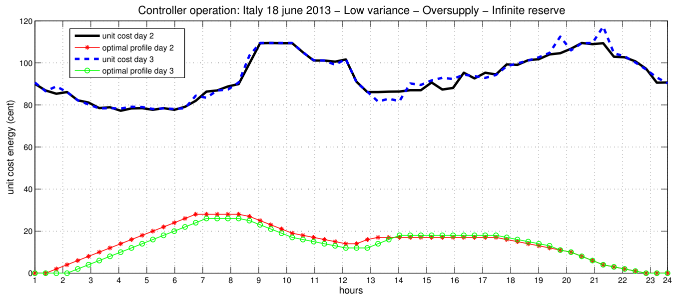

Figure 47: Italy, 18 june 2013 - low variance - oversupply - infinite reserve: planning phase.

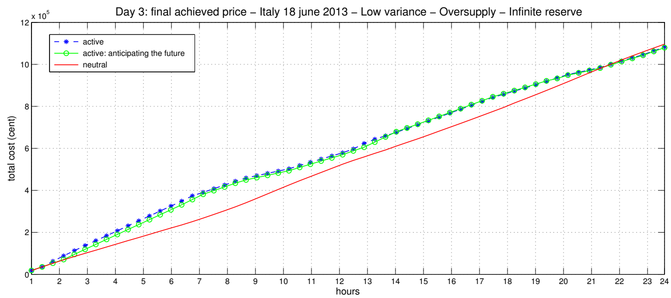

Figure 48: Italy, 18 june 2013 - low variance - oversupply - infinite reserve: active vs neutral.

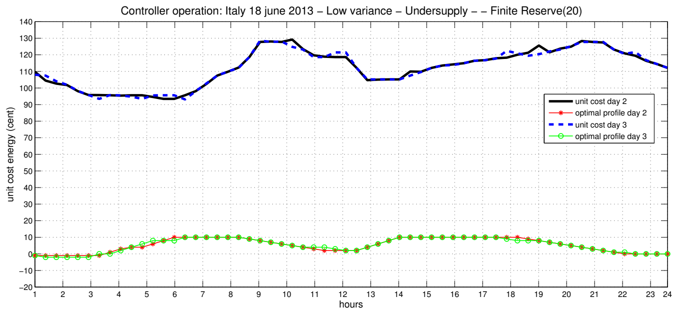

Figure 49: Italy, 18 june 2013 - low variance - undersupply - finite reserve: planning phase.

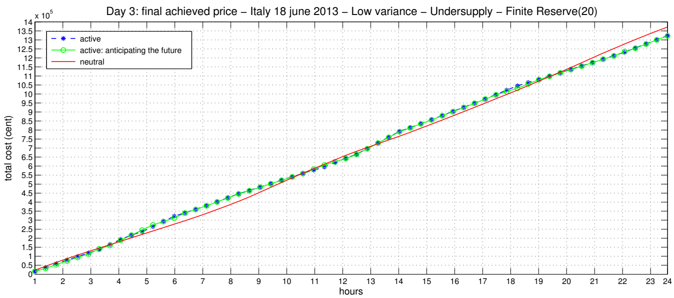

Figure 50: Italy, 18 june 2013 - low variance - undersupply - finite reserve: active vs neutral.

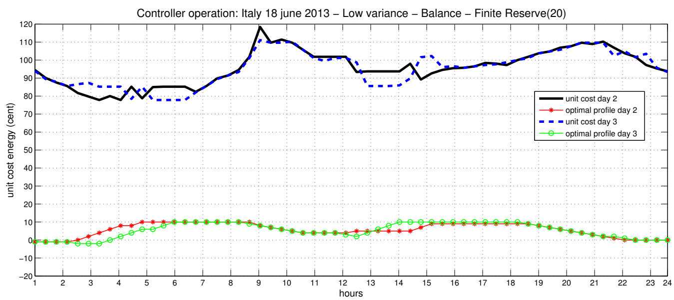

Figure 51: Italy, 18 june 2013 - low variance - balance - finite reserve: planning phase.

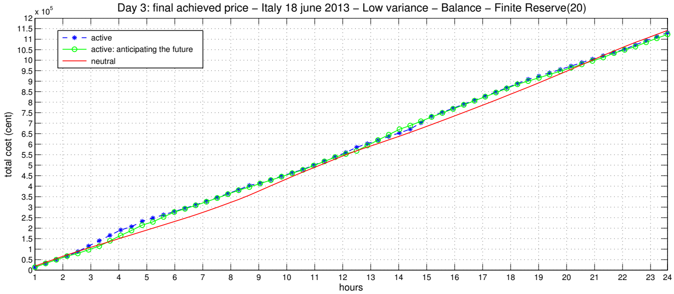

Figure 52: Italy, 18 june 2013 - low variance - balance - finite reserve: active vs neutral.

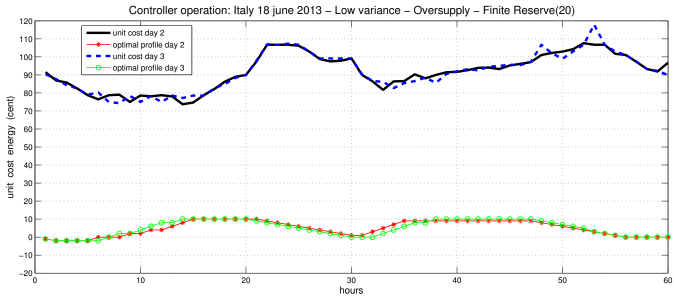

Figure 53: Italy, 18 june 2013 - low variance - oversupply - finite reserve: planning phase.

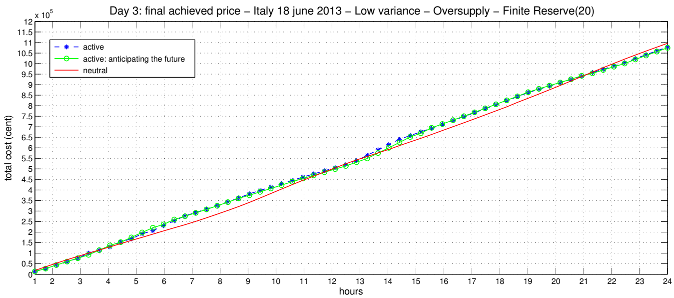

Figure 54: Italy, 18 june 2013 - low variance - oversupply - finite reserve: active vs neutral.

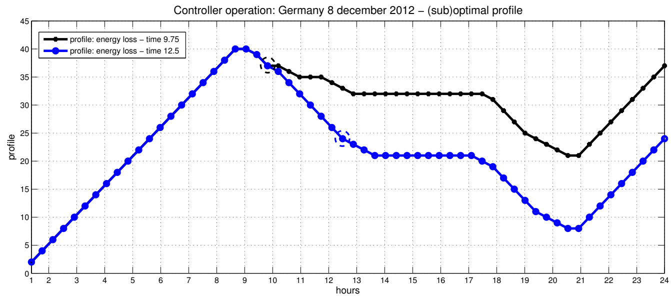

Figure 55: Controller operation: - Germany 8 december 2012 - Average variance - Undersupply - Infinite reserve - (sub)optimal profile.

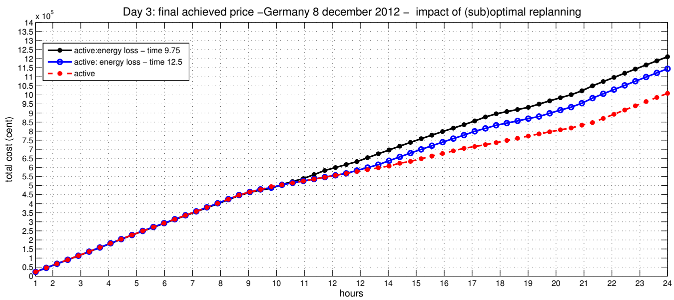

Figure 56: Day 3: final achieved price - Germany 8 december 2012 - Average variance - Undersupply - Infinite reserve - Impact of (sub)optimal replanning.

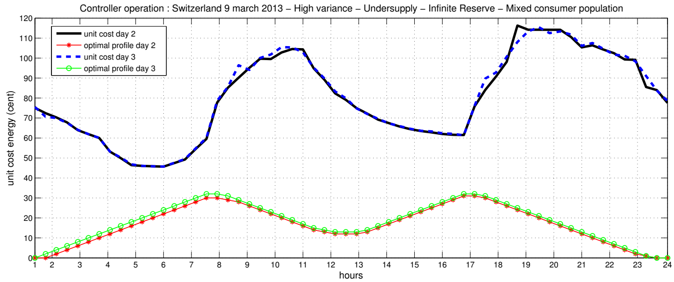

Figure 57: Switzerland, 9 march 2013 - high variance - undersupply - infinite reserve - mixed consumer population: planning phase.

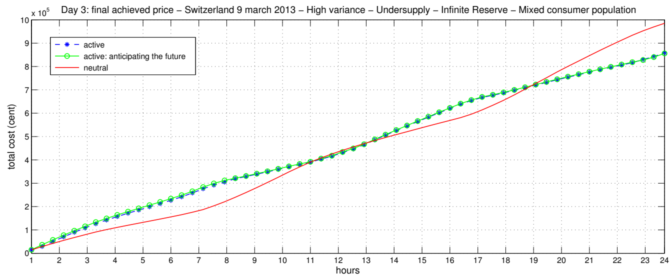

Figure 58: Switzerland, 9 march 2013 - high variance - undersupply - infinite reserve - mixed consumer population: active vs neutral.

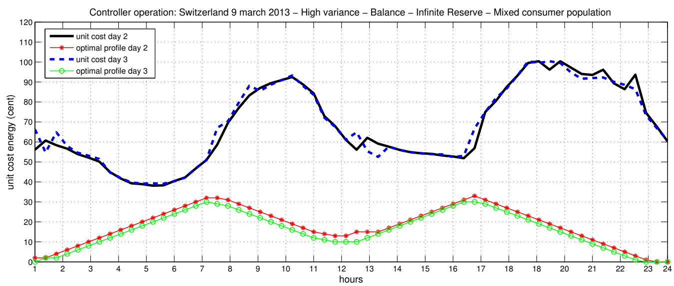

Figure 59: Switzerland, 9 march 2013 - high variance - balance - infinite reserve - mixed consumer population : planning phase.

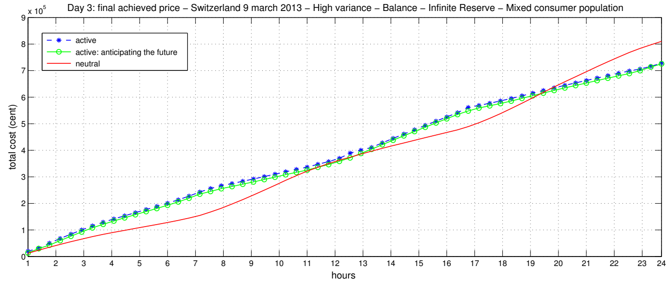

Figure 60: Switzerland, 9 march 2013 - high variance - balance - infinite reserve - mixed consumer population: active vs neutral.

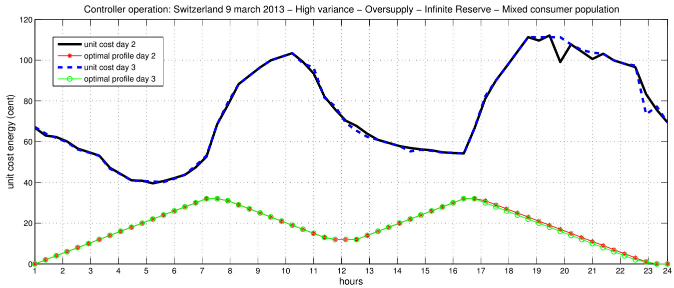

Figure 61: Switzerland, 9 march 2013 - high variance - oversupply - infinite reserve - mixed consumer population: planning phase.

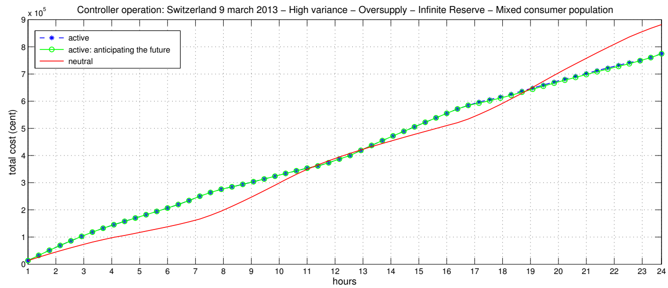

Figure 62: Switzerland, 9 march 2013 - high variance - oversupply - infinite reserve - mixed consumer population: active vs neutral.

Chapter 7

Conclusions and Future Work

In this thesis we have proposed a controller based on which end-users may optimize their energy consumption and better manage their own electricity costs. The idea behind our approach is to make elastic the demand for, and the supply of, electricity of prosumers in order to optimize their energy cost based on power market conditions and on suitable constraints on their power consumption. We focus, in particular, on the planning phase of the consumption [42] of prosumers based on the negotiation mechanism of DEZENT. The general idea behind the planning (optimization) phase of a consumer is to plan consumption as smartly (delay or anticipate the consumption) as possible during a day, a week, a month or even a year. The results of this behavior is the minimization of the electricity cost in the long run, under certain assumptions on energy costs, as resulted by local negotiations, and on the acceptable variations of consumer requests.

The approach we proposed depends on the following parameters: (a) the profile of the free market power cost in every slot, and in particular its variance; (b) the prosumer environment: heavy production, heavy consumption or equilibrated prosumer population; (c) how to estimate the cost of energy after the DEZENT negotiation; (d) the amount of allowed power increase/decrease; and (e) the available reserve. Items (a) and (b) affect the simulations but their values are not directly available to the prosumer, while items (c), (d) and (e) are available inputs for prosumers optimal planning.

We have performed extended experimental studies based on the available DEZENT simulator and on a Java implementation of the optimal controller. An experiment depends essentially on three parameters: (1) the free market power cost, which can exhibit high, low or average variance: for this we chose real data from the day ahead market prices of Switzerland, Italy and Germany respectively; (2) the prosumer environment, namely heavy production, heavy consumption or equilibrated prosumer population; and (3) the available reserve capacity which can be either finite or infinite. For every combination of the parameters we synthesized the optimal controller and computed the gain with respect to a neutral consumer operating in the same environment (i.e. within a context of other prosumers, which are the same in both cases). As expected, the highest energy cost reduction was obtained when we had a high variance on the profile cost of the electricity, the prosumer environment was in the undersupply situation and the reserve capacity of the controller was infinite. Vice versa, the lowest energy cost reduction has corresponded to low variance, oversupply situation and finite reserve capacity.

Furthermore, the study of the problem of (sub) optimal repeated replanning for the rest of the day has shown that a prosumer having to consume more than expected will pay a remarkable additional cost at the end of the day which depends also on the increased unit costs; in the case in which a part of the available energy reserve is lost, the additional cost paid is proportional to the amount of energy lost. However, replanning is definitely convenient with respect to continuing with the old plan, just adapted for the loss of power.

A main contribution of the thesis is the combination of the reinforcement learning mechanism at the DEZENT level and of the control mechanism used for profile optimization. These two mechanisms are not independent: rather, a negative feedback loop is present, between them.

During the reinforcement learning step the agent updates the weights of its negotiation strategies according to the amount of energy it needs. Next, the control mechanism takes into account the energy cost which is the effect of the negotiation to determine how much energy it requires. The simulations already show the effects of this feedback loop.

One of the open issues is the global feedback effect of our introduced controller in a consumer population where each consumer can make use of it. In fact each controller in use might, independently and locally, increase its consumption during periods in which the electricity price turns out to be low. Hence at the prosumer population level, the electricity demand will increase, possibly quite a lot, during those periods. This feedback effect will lead to the increase of the energy cost during periods in which the price is in reality convenient. Eventually the system may oscillate. This may lead to congestion issues on the behavior of the controller.

For the study of the congestion issue, additional simulations have been made where more than one active consumer is present in the environment. The possible congestion problem was then made explicit by the negative feedback effect already mentioned above, now increased due to the multiplicity of active prosumers. Again the main result we obtain is that the agent has spent less when making used of the control mechanism (see Figures 58, 60 and 62). The comparison between an active and a neutral prosumer was still meaningful, since it shows, in the Nash style, the convenience for a single prosumer to adopt the active strategy. Of course more extensive simulations would be needed to study the congestion problem (evidenced e.g. by a difficult convergence of the learning process), but on the one side we did not find any hint of congestion in our limited experiment, on the other hand the literature on reinforcement learning in the presence of congestion (e.g. the El Farol problem [18, 39], described in Section 2.1) could suggest convenient countermeasures easily applicable in the DEZENT approach.

Future work might consider particular kinds of prosumers and their energy storing devices, e.g. batteries of electric cars, modeling their peculiar abilities of adapting their needs to variable energy cost profiles.

References

[1] IEEE Smart Grid World Forum Brussels, Belgium 2-3 December 2010. 1

[2] Smart Grid technology Conference & EXPO, San Diego California 2-3 june 2010. 1

[3] http://smartgrid.ieee.org/nist-smartgrid-framework. 1

[4] ETP Smartgrid & SAP. 1

[5] http://www.smartgrids.eu. 1

[6] IEC 61850 Communication Networks and Systems in substantions. 2

[7] http://www.twenties-project.eu. 3

[8] http://psychology.about.com/od/behavioralpsychology/Behavioral Psychology.htm. 14

[9] http://en.wikipedia.org/wiki/Electricity_market. 22

[10] http://en.wikipedia.org/wiki/Virtual_Bidding. 29

[11] Energy Conservation Committee Report and Recommendations, Reducing Electricity Consumption in Houses. Ontario Home Builders' Association, May 2006. 46

[12] The 2nd IEEE INFOCOM on Communications and Control for Smart Energy Systems (CCSES 2013). 54

[13] European Power Exchange. http://www.epexspot.com. 54

[14] http://www.e-energy.de. 54

[15] Gestore Mercati Elettrici. http://www.mercatoelettrico.org. 54

[16] European Energy Exchange: EEX Home page. http://www.eex.com. 54

[17] Biancardi Alberto and Fontini Fulvio. *Liberi di scegliere? Mercati e regole nei settori dell'energia*. Il Mulino, 2005. 22

[18] W Brian Arthur. Inductive reasoning and bounded rationality. *American Economic Review*, 84(2):406–411, 1994. 18, 79

[19] L.A. Barroso, T.H. Cavalcanti, P. Giesbertz, and K. Purchala. Classification of electricity market models worldwide. *CIGRE/IEEE PES, International Symposium*, pages 9–16, 2005. 6, 25, 29

[20] Carsten Block, John Collins, and Wolf Ketter. Agent-based competitive simulation: Exploring future retail energy markets. In *Proceedings of the 12th International Conference on Electronic Commerce (ICEC)*, Hawaii, USA, August 2010. 26

[21] Severin Borenstein, James Bushnell, and Frank Wolak. Measuring market inefficiencies in california's restructured wholesale electricity market. *The American Economic Review 92*, 5:1376 – 1405, 2002. 22

[22] A. Botterud, J. Wang, and C. Monteiro. Wind power forecasting and electricity market operations. In *Proceedings of USAEE*, 2009. 23

[23] D. Challet and Y.-C. Zhang. Emergence of cooperation and organization in an evolutionary game. *Physica A: Statistical Mechanics and its Applications*, 246(34):407 – 418, 1997. 20

[24] H. Chao, S. Peck, S. Oren, and R. Wilson. Flow-based transmission rights and congestion management. *Electricity Journal*, pages 38–58, October 2000. 26

[25] Peter Cramton. Competitive bidding behavior in uniform-price auction markets. *Published in the Proceedings of the Hawaii International Conference on System Sciences*, January. 29

[26] Berry D.A. and Fristedt B. *Bandit Problems: Sequential Allocation of Experiments*. Monographs on Statistics and Applied Probability, Chapman and Hall, London/New York, 1985. 15

[27] Ido Erev and Alvin E. Roth. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, 88(4):848–881, 1998. 18

[28] Robert Gibbons. *Game Theory for Applied Economists*. Princeton University Press, Princeton, New Jersey, 1992. 14

[29] J. Gordijn and J. M. Akkermans. Business Models for Distributed Energy Resources In a Liberalized Market Environment. In *Electrical Power Systems Research Journal*, volume 77, pages 1178–1188. 22

[30] John C. Harsanyi. Games with incomplete information played by "bayesian" players, i-iii. part i. the basic model the basic model. *Management Science*, 14(3):159–182, November 1967. 22

[31] G. Hart. Nonintrusive appliance load monitoring. *In Proc. of the IEEE*, pages 1870–1891, 1992. 47

[32] N. Hatziargyriou, H. Asano, R. Iravani, and Marnay. An overview of ongoing research, development, and demonstration projects. In *Berkeley Lab Publications LBNL-62937, Lawrence Berkeley National Laboratory*, July 2007. 23

[33] Joskow and Kahn. A quantitative analysis of pricing behavior in California's wholesale electricity market during summer 2000 . In *NBER Working Paper Series 8157, National Bureau of Economic Research*, 2001. 22

[34] G. Kariniotakis, I. Waldl, I. Marti, G. Giebel, T. Nielsen, J. Tambke, J. Usaola, F. Dierich, A.Bocquet, and S. Virlot. Next generation forecasting tools for the optimal management of wind generation . In *Probabilistic Methods Applied to Power Systems, International Conference*, pages 1–6, 2009. 23

[35] R. Klempka and B. Swiatek. In Day-Ahead electricity load forecasting. In *Electrical Power Quality and Utilisation, 10th International Conference on*, pages 1–5, 2009. 26

[36] Koen Kok, Cor Warmer, and Ren Kamphui. The PowerMatcher: Multiagent control of electricity demand and supply. In *IEEE Intelligent Systems*, volume 21, pages 89–90. Part of overview article: "Agents in Industry: The Best from the AAMAS 2005 Industry Track". 23

[37] Koutsopoulos and L. Tassiulas. Challenges in demand load control for the smart grid. *Energy Efficient Computing and Networking*, 2011. 46

[38] Koutsopoulos and L. Tassiulas. Control and optimization meet the smart power grid: Scheduling of power demands for optimal energy management. *Proc. Intl. Conf. Energy Efficient Computing and Networking*, 2011. 46

[39] Andrew Moore Leslie Pack Kaelbling, Michael Littman. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285, 1996. 14, 15, 16, 79

[40] Matteo Marsili and Andrea De Martino. Statistical mechanics of socio-economic systems with heterogeneous agents. Working papers, Warwick Business School, Financial Econometrics Research Centre, 2006. 20

[41] Albert Molderink, Vincent Bakker, Maurice G.C. Bosman, Johann L. Hurink, and Gerard J.M. Smit. Management and control of domestic smart grid technology. *IEEE Transactions on Smart Grid*, 1(2):109–119, September 2010. 46

[42] Ugo Montanari and Alain Tcheukam. Real time market models and prosumer profiling. *IEEE Conference on Computer Communication Workshop*, pages 7–12, April 2013. Turin, Italy. 6, 77

[43] Williams R.J and Baird L.C. Analysis of some incremental variants of policy iteration: First steps toward understanding actor-critic learning systems. Technical report, Technical Report NU-CCS-93-11 Northeastern University, college of Computer Science, Boston, Ma 02115, September 1993. 17

[44] Robert Rosenthal. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory*, 2(1):65–67, 1973. 18

[45] A. Rosin, T. Moller, M. Lehtla, and H. Hoimoja. Analysis of household electricity consumption patterns and economy of water heating shifting and saving bulbs. *Scientific Journal of RTU*, 1(2):15–20, September 2010. 46

[46] Steven Stoft. *Power System Economics: Designing Markets for Electricity*. Wiley-IEEE Press, 2002. 6

[47] Richard Sutton, Andrew Barto, and Charles Anderson. *Reinforcement Learning: An Introduction.* The MIT Press, 1983. 16

[48] Richard S. Sutton. Learning to predict by the methods of temporal differences. *Machine Learning*, 3:9–44, 1988. 16, 17, 42

[49] H.F. Wedde, S. Lehnhoff, C. Rehtanz, and O.Krause. A distributed real-time multi-agent system for decentralized management of electric power. *18th Euromicro Conference on Real-Time Systems, Dresden, Germany*, 2006. 6, 7, 33, 34

[50] H.F. Wedde, S. Lehnhoff, C. Rehtanz, and O.Krause. Autonomous real-time management of unpredictable power needs and supply. *Technical Report #810, University of Dortmund, Germany*, 2007. 6, 7, 33, 34

[51] H.F. Wedde, S. Lehnhoff, C. Rehtanz, and O.Krause. Bottom-up self-organization of unpredictable demand and supply under decentralized power management. *Self-Adaptive and Self-Organizing Systems, Second IEEE International Conference*, pages 74–83, 2008. 7, 33, 34, 42

[52] H.F. Wedde, S. Lehnhoff, C. Rehtanz, and O.Krause. Distributed learning strategies for collaborative agents in adaptive decentralized power systems. *15th Annual IEEE International Conference and Workshop on the Engineering of Computer Based Systems*, pages 26–35, 2008. 7, 33, 34

[53] Duncan Whitehead. The el farol bar problem revisited: Reinforcement learning in a potential game. ESE Discussion Papers 186, Edinburgh School of Economics, University of Edinburgh, September 2008. 19

[54] Robert Wilson. Architecture of power market. *Econometrica*, pages 1299–1340, 2002. 6, 21