Automatic Tariff Generation for Electricity Markets Using Reinforcement Learning

by

Jonathan Serrano Cuevas

Submited in partial fulfillment of the requirements for the degree of

Master in Computer Science

Instituto Nacional de Astrofísica, Óptica y Electrónica
September, 2014
Tonantzintla, Puebla

Supervisor:

Jose Enrique Muñoz de Cote
Computer Science Department
INAOE

To my parents and family, for all their support.

Abstract

The most widely used energy sources rely on oil and therefore are not going to supply energy to industrial processes and people's lives forever. Most of the electric energy produced nowadays comes from thermoelectrical plants, which are industrial complexes with a furnace at its heart burning oil to generate steam. The large amounts of electrical energy is then transmitted over the electrical grid to factories, offices and houses. This centralized scheme has several disadvantages. The first one is its oil dependency, because the oil's price increases constantly. The second one is the transmission grid by itself. Furthermore, this grid is very old, and because of its nature, a localized problem has the potential to start a snowball effect massive blackout. An alternative to address both the dependency on oil of current energy generation and its centralized nature is a technology called smartgrid. Smartgrid enables an electrical network with communications, sensing, self-healing and digital capabilities. These features transform the old, centralized energy network into a distributed network, which can integrate small producers and consumers (or hybrids that can be both, consumers and producers) on almost any point of the network at any time. With this scheme, large power suppliers will have its place on the network, but also an increasing amount of small green-energy producers, which by aggregation, will supply large amounts of power to consumers, creating a liberalized energy market. The idea of this market is for it to be driven by brokers, which are entities that can buy energy from producers and sell it to consumers, by means of contracts called *tariffs*, which are regulated by constraints set by a regulator entity.

These constraints have the purpose of ensuring the health of the electric grid, while providing the energetic needs to the population being served.

This thesis work proposes an autonomous learning agent that is capable of perceiving the market needs and constraints in order to create tariffs to be offered to consuming customers (energy producers and/or consumers). This agent is the tariff expert part of a complete broker called COLD broker, which works on specific aspects of the smart grid (see section 3.1 for details on this). From now on, when the term COLD broker is used it will refer to the tariff expert part of the agent, unless explicitly stated otherwise.

One of the contributions of this thesis is a market representation, which is independent to the number of brokers and the number of tariffs available, and a set of market-bounded actions. The number of market states will not grow as the number of brokers or tariffs increases, making it a very scalable representation. The market representation and the set of actions are used by COLD broker to create an MDP and learn a policy that will provide it with the best possible profit on the long run (see section 4), by choosing an appropriate pricing scheme for consumption and production tariffs. This profit depends on the prices and the amount of the energy bought and sold, and on the tariffs the competing agents offer. In order to analyze the market and to have a testing framework, Power TAC was used. Power TAC is a complex smartgrid simulator, which accounts for fairly real aspects of what might be an actual liberalized energy market.

COLD broker was tested against fixed-strategy brokers and against a learning broker proposed by recent investigation works. Average and standard deviations were measured on each experiment to determine which broker's behaviour was consistently better. On average, COLD broker's utility roughly doubled its closest competitor's utility.

Resumen

Las fuentes de energía más usadas se basan principalmente en el uso de petróleo, y por esta razón, no podrán satisfacer las necesidades de los procesos industriales y de las vidas de las personas para siempre. La mayor parte de la energía eléctrica producida hoy en día proviene de plantas termoeléctricas, que son complejos industriales con un proceso de combustión que quema petróleo, o alguno de sus derivados, para generar vapor. La energía generada por estos procesos es entonces transmitida a través de la red eléctrica hasta las fábricas, oficinas y casas que la requieren. Este esquema centralizado tiene varias desventajas. La primera es su dependencia del petróleo, debido a que su precio aumenta constantemente. El segundo es la red de transmisión por si misma, debido a que por su naturaleza, un problema localizado tiene el potencial de comenzar un fallo generalizado a manera de avalancha. Una alternativa para mitigar la dependencia al petróleo y la naturaleza centralizada de la red de distribución de energía eléctrica actual es una tecnología llamada smartgrid. Smartgrid puede proveer a la red eléctrica de habilidades de comunicación, sensado y auto-reparación. Estas habilidades transformarían la red de transmisión centralizada en una red distribuida, capaz de integrar pequeos productores y consumidores (o modelos híbridos que pueden ser tanto consumidores como productores) en cualquier punto de la red y en cualquier momento. Con este esquema, las grandes centrales productoras de energía tendrían su lugar en la red, pero también lo tendrían un creciente número de pequeos productores quienes, en su conjunto, podrían proveer de enormes cantidades de energía eléctrica a los consumidores a través de un mer-

cado abierto. Este mercado sería manejado por brokers, que son entidades capaces de comprar energía de los productores para venderla a los consumidores, a través de contratos llamados tarifas. Las características de estas tarifas estarían reguladas por un organismo central, con el objetivo de asegurar la estabilidad de la red eléctrica.

Esta tesis propone un agente autónomo con capacidades de aprendizaje, que tiene la habilidad de percibir las necesidades del mercado con el objetivo de crear tarifas para los clientes (productores o consumidores de energía). Este agente es el experto en tarifas de un agente más complejo llamado COLD broker, capaz de desenvolverse en otros aspectos de un mercado de energía abierto (ver sección 3.1 para encontrar más detalles sobre este tema. De ahora en adelante, el término COLD broker se referirá al experto en tarifas del agente completo, a no ser que se especifique cualquier otra cosa.

Unas de las contribuciones de esta tesis es una representación de un mercado de energía, que es independiente del número de brokers y del número de tarifas disponibles; así como un conjunto de acciones acotadas a los precios del mercado. En la representación propuesta, el número de estados no va a aumentar cuando lo hagan el número de brokers o de tarifas, lo que la vuelve muy escalable. La representación del mercado y el conjunto de acciones son usados por COLD broker para crear un MDP y aprender una política capaz de proporcionarle la mejor utilidad posible en el largo plazo (ver sección 4). Esto lo logra escogiendo esquemas de precios adecuados para crear tarifas de consumo y producción. Con el objetivo de analizar el mercado se usó PowerTAC como plataforma de pruebas, que es un simulador muy completo de mercados de energía abiertos.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Traditional energy sources that have satisfied human energy requirements during the last century are going to be depleted in the near future. This cannot happen before finding sustainable, robust and manageable alternative energy sources; otherwise severe problems will emerge damaging almost every aspect of human life. So far there are many alternatives to outline, such as wind or solar energy. Some other technologies are being explored that might produce energy from rivers or even the tides [Ben Elghali et al., 2007]. However, all these eco-friendly energy sources have a major drawback: they are not predictable and thus, it is not possible to fully rely only on them to satisfy vital energy-consuming processes. There has to be some regulating entity or entities that are capable of matching energy consumption and production to provide a robust system that will minimize blackouts or energy supply issues.

This is where smartgrid technologies are becoming more and more useful. The smartgrid relies on the power supply network that interconnects each and every energy producer and consumer. This technology uses the power wires to broadcast and receive information to act upon it, with the objective of improving the efficiency, reliability, economics and sustainability of the production, consumption and distribution of electrical energy [Ipakchi and Albuyeh, 2009].

Smartgrid technologies enable producers to sell energy to consumers by using

a broker as an intermediary. This scenario allows a model where hundred, or even millions, of small producers can match the energy production capacity of large centralized energy plants such as thermoelectrical plants, hydroelectical dams or even nuclear plants. However, this freedom introduces new obstacles, which if not handled properly, can lead to blackouts and other severe problems. Such issues might be related to brokers executing bad trading practices. To handle these obstacles, a set of rules and regulations are to be followed by all participants, producers and consumers, to guarantee a robust flow of energy which satisfies a variable demand.

The work to be done now is to design and improve the game rules so as to have an efficient market, but before that, a comprehensive study has to take place in order to understand the dynamics of a descentralized energy market, where complex components such as the weather and the human behaviour, affect this dynamics. To work on this problem platforms such as Power TAC [Ketter and Collins, 2013] have been developed. Power TAC simulates an entire energy market with producers, consumers and brokers buying and supplying energy. Power TAC allows us to study energy markets in a way that has not been done before. Power TAC is a complex simulation environment that consists of several components. In this thesis we focus on designing an expert broker on tariff markets.

On the design of a tariff expert several aspects are to be considered. First we have to realize that there are many types of customers with their own set of preferences. We also need a way to assess how successful or unsuccessful a tariff is to be able to improve it. We must as well consider other brokers tariffs characteristics. There is also a need to obtain profit with the offered tariffs, otherwise the broker model will not be sustainable. All these aspects were taken into account to design our tariff expert broker. All the latter issues are closely related, because the customers choose among the available tariffs the one that yields them the lower cost in terms of their preferences. If a broker publishes tariffs which fit customer's preferences, then it will obtain a utility. The utility function stated on Sec. 4.2 fully considers the

income obtained and the expenses incurred, addressing the profit issue. This utility function considers the income of selling energy. It also considers both the expenses of buying this energy and the costs of generating an imbalance. On the other hand, the market representation explained on Sec. 4 accounts for the other brokers actions. This representation is an important contribution of this thesis work, because its complexity does not increase neither with the number of competing brokers nor with the number of published tariffs, making it very scalable. Lastly, the COLD broker implementaion carefully tracks the performance of each published tariff, in order to provide enough information to feed the MDP.

1.1 OBJECTIVES

General Objective: design and test a broker that uses reinforcement learning to generate electric energy tariffs, accounting for other brokers tariff, with the purpose of maximizing its own utility in the long term.

Specific Objectives:

1. Design a state representation to facilitate the learning and decision making processes.
2. Design and test a set of market actions and strategies which allow the broker to react appropriately to other broker's actions.
3. Learn using reinforcement learning (RL) the environment dynamics so as to maximize profit.

1.2 CONTRIBUTIONS

This thesis work developed a tariff expert broker capable of learning how to attract customers and maximize profits in the long run. Our market representation and set of actions (strategies) are novel and, as will be shown in Sec. 4.2.2, they allow the learning broker to adapt to fast changing and non-stationary environments. We designed a representation of the tariff energy market whose size, as mentioned before, remains constant with the number of competitors (*i.e.* scales gracefully); and that compactly encodes the required information for decision making. Also, the set of actions allows the broker to transition easily (and fast) across states, which translates into attractive price-based tariffs to customers. The specific contributions of this work are the following ones:

- Proposed an adaptable, broker-independent market representation.
- Designed an offline learning method capable of transfering the learned knowledge.
- Developed a learning broker designed to create tariffs for consumption and production consumers.

1.3 THESIS OUTLINE

Chapter 1 states the main and specific objectives of this thesis work and also lists its contributions. Chapter 3 describes the related work on smart grid, energy tariffs and automatic tariff generation. Chapter 2 provides an insight on the theory required to understand the broker develpment and test procedure, including MDP and reinforcement learning. On this chapter a section is dedicated to Q-Learning, because this was the selected method to learn the required policy which maximizes

profit for the broker. This chapter includes as well a comprehensive explanation of the deregulated energy market simulator Power TAC. The explanation describes the general structure of the platform and also the customer models, and how they evaluate a tariff belonging to any given broker. Chapter 4 describes the broker's design process by clearly stating its objective, actions and market representation. The broker's MDP is described as well on this section. Chapter 5 describes the experimental setup which tested the broker against various scenarios. This section also describes each experiment and provides a discussion for each of them. Finally, the last two chapters include the conclusions derived from this thesis work and the future work that could improve the results obtained. A glossary is included as well at the end of the document.

Chapter 2

Background

This section describes the main theory concepts required to develop the tariff expert broker. The broker models the environment as a Markov decision process (MDP) and uses concepts of machine learning to learn a policy directly from interaction with the environment provided by the simulator. Therefore this section includes backgrounds on types and classification of tariffs (Sec. 2.1), MDPs (Sec. 2.2) machine learning (Sec. 2.3) and the simulator environment (Sec. 2.4).

2.1 Tariffs in the SmartGrid

In the literature several types of tariffs for energy market exist. In general most tariff have two main cost components: the electricity commodity and a risk premium paid to protect customers against price variations [Ilie et al., 2007]. The way on which these two costs components are structured depends on the the type of tariffs. On this section we discuss the most common tariff schemes.

2.1.1 Time Independent Tariffs

As their name imply, these tariffs have rates that do not depend on the time of usage.

FLAT TARIFFS

A flat tariff is the most simple tariff. In a flat tariff, customers pay a fixed price per KWh, with no dependency at all on the time the energy is consumed. A flat tariff has high risk premiums and it provides a large benefit to customers that have a high consumption during peak hours [Faruqui, 2010], because these customers will not have to pay an extra fee during these hours with energy scarcity, which means that brokers have to buy their energy at high spot prices at the wholesale market.

BLOCK TARIFFS

Another time-independent tariff is the increasing block tariff. In this tariff scheme, customers pay a fixed price until a certain consumption threshold is reached after which the customers have to pay a higher rate [Borenstein, 2009]. There might be multiple thresholds in the tariff, this means that the consumption rates might increase several times. The risk premium of this tariff is directly associated with the amount of energy consumed by customers, which means that as consumers increase their consumption, they have to pay larger amounts of money at the risk of having higher prices. The broker's risk also grows as consumers require more energy.

2.1.2 TIME DEPENDENT TARIFFS

Time dependent tariff schemes charge customers according to a function that considers the volume of energy consumed and the time of the consumption. A common reference for this time of consumption are the peak hours.

TIME-OF-USE TARIFFS

On this scheme the customer pays a higher rate during the peak hours of a day and lower prices during off-peak hours; thus consumers have an incentive to use high

energy demanding devices during off-peak hours, which reduce the consumption during peak hours. Consumers that have a higher demand during off-peak hours or consumers that can easily switch their consumption to other times can benefit widely from using time-of-use tariffs [Kirschen, 2003].

Critical-peak pricing

Critical-peak pricing is actually a combination of flat tariffs and time-of-use tariffs. A critical-peak pricing scheme charges customers an extremely high price under certain predetermined conditions, considered critical. An example of a critical condition might be extremely high or low temperatures, where high energy-demanding devices might be used to control indoor temperatures. The risk of price changes during this unusual events is thus fully transferred to the customers by using this tariff scheme [Borenstein et al., 2002].

Real-time pricing

A Real-time pricing tariff is a tariff in which customers are charged with a rate, on a hourly basis, that may vary each day. The rates for each day can be determined 24 hours in advance or earlier [Barbose et al., 2004]. Consumers then will be charged by a rate that adapts to the actual energy's market prices. By doing this, consumers will pay more money when the overall consumption is high to incentivize using less energy. Expert economists state that real-time pricing tariffs are considered to be the most efficient approach in reducing peak demands by using price incentives [Borenstein et al., 2002]. On this tariff scheme, many risks are transferred to the customer, since price changes on the wholesale market can be predicted easier, on a short term, and therefore the prices in real-time pricing tariffs are closer to the prices on the wholesale market.

GREEN TARIFFS

The energy traded via green tariffs is produced in an enviromentally friendly way. Green energy usually is more expensive to produce, compared to energy produced using oil, so customers have to pay an extra premium. However, green tariffs might be attractive to certain type of customers that perceive benefits from using green energy [Haas et al., 2011].

2.2 MARKOV DECISION PROCESS (MDP)

A Markov decision process, or MDP, provides a mathematical framework for decision making in environments where outcomes are partly random and partly under the control of a decision maker or broker[Puterman, 2005]. More precisely, a MPD is a discrete time stochastic control process which can be defined as a tuple:

$$M = \langle S, A, P, R \rangle \,, \tag{2.1}$$

where:

- S is a finite set of states,
- A is a finite set of actions
- P is the probability that action $a \in A$ in state $s \in S$ at time t will lead to state s' at time $t + 1$,
- R is the immediate reward received after transition to state s' from state s.

The tuple represents the MDP at any given time and it is evaluated and updated by the decision maker at time steps. At each time step, the process is in some

state $s \in S$, and the decision maker may choose any action $a \in A$ that is available in state s. The process responds at the next time step by randomly moving into a new state $s' \in S$, and giving the decision maker a corresponding reward R.

The probability that the process moves into its new state s' is influenced by the chosen action. Specifically, it is given by the state transition function $p_t(s'|s, a)$. Thus, the next state s' depends on the current state s and the broker's action a. But given s and a, the transition is conditionally independent of all previous states and actions; in other words, the state transitions of an MDP possess the Markov property. A stochastic process has the Markov property if the conditional probability distribution of future states of the process depends only upon the present state, and not on the sequence of events that preceded it.

MDP's are widely used on several fields in order to determine which decision is the best at any given time over a certain environment state. Making the correct decision has both immediate and long term consequences, so this decisions are not meant to be taken into isolation. An MDP provides a sequential decision model that assigns a numeric value to the consequences of picking any of the available actions, and it is sufficiently broad to allow modeling most realistic sequential decision-making problems. An agent can use an MDP with the ultimate goal of choosing a sequence of actions at every decision step, which causes the system to perform optimally with respect to some predetermined performance criterion [Puterman, 2005]. The current section briefly describes an MDP states and actions, and how these two elements relate to each other by transision probabilities and rewards.

2.2.1 Rewards and Transition Probabilities

As a result of choosing action $a \in A$ in state s at decision step t, the agent will receive a reward $R_t(s, a)$ and the system state s' at the next decision step is determined by the probability distribution $P_t(s'|s, a)$.

The reward R_t might be:

- a lump sum received at a fixed or random time prior to the next decision step,
- accrued continuously throughout the current decision step and the next one,
- a random quantity that depends on the system state at the subsequent decision step,
- or a combination of any of these options.

The reward received by the learning broker developed on this thesis is accumulated after each decision step and before the next one in the form of earnings received by trading electric energy.

2.2.2 POLICIES

Given a state, the agent has to decide which action is to be executed, so a policy prescribes a procedure for action selection in each state at a specified decision step. Policies range in generality from deterministic markovian to randomized history dependent, depending on how they incorporate past information and how they select actions. This thesis develops a deterministic markovian policy. These type of rules are functions $d_t : S \rightarrow A_s$, which specify the action choice when the system occupies state s at decision point t. For each $s \in S$, $d_t(s) \in A$. This policy is said to be Markovian because it depends on previous system states and actions only through the current state of the system, and deterministic because it chooses an action with certainty. A policy is said to be stationary if $P(s) = a$ for all $t \in T$ and all $s \in S$.

As an example, the formal description of the MDP on Fig. 2.1 is as follows:

- Decision steps: $T = \{1, 2, ..., N\}, N \leq \infty$

- States: $S = \{s_1, s_2\}$
- Actions: $A_{s_1} = \{a_{1,1}, a_{1,2}\}, A_{s_2} = \{a_{2,1}\}$
- Rewards: $R_t(s_1, a_{1,1}) = 5,\ R_t(s_2, a_{2,1}) = -1,\ R_t(s_2, a_{2,1}) = -1$
- Transition probabilities:

$$\begin{aligned} P_t(s_1|s_1, a_{1,1}) &= 0.5 \quad & P_t(s_2|s_1, a_{1,1}) &= 0.5 \\ P_t(s_1|s_1, a_{1,2}) &= 0 \quad & P_t(s_2|s_1, a_{1,2}) &= 1 \\ P_t(s_1|s_2, a_{2,1}) &= 0 \quad & P_t(s_2|s_2, a_{2,1}) &= 1 \end{aligned}$$

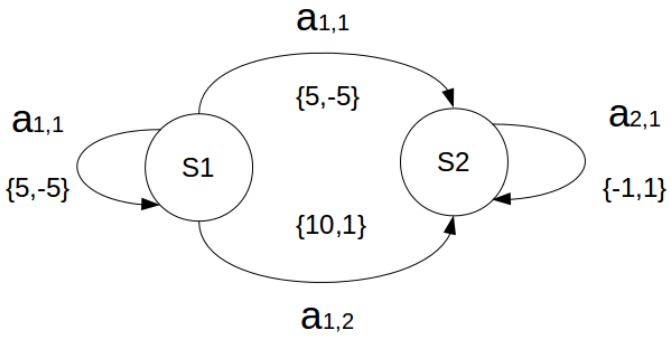

Figure 2.1: An MDP example

2.3 Reinforcement Learning

According to [Barto, 1998], reinforcement learning (RL) is the process of learning how to map situations to actions, so a numeric reward signal is maximized. The learner is not told which actions to take, as in most forms of machine learning, but instead must discover which actions yield the most reward by experimenting with them. In the most interesting and challenging cases, actions may affect not only the immediate reward but also the next situation and, through that, all subsequent rewards. These two characteristics, trial-and-error search and delayed reward, are the two most important distinguishing features of RL.

Supervised learning is different from reinforcement learning because the first one requires examples provided by a knowledgeable supervisor while the latter learns from interaction with its environment. Facing this interaction, one of the challenges that arise is the trade-off between exploration and exploitation. A trained learner might know what action to execute given any state to obtain a large reward, however, to learn that, the learner had to try all the possible actions available on that specific state; and some of those actions may yield low rewards, even negative ones. To maximize its utility the agent must try a variety of actions and progressively favor those that appear to be best. On a stochastic task, each action must be tried many times to gain a reliable estimate of its expected reward.

2.3.1 Reinforcement Learning Elements

Besides the agent and the environment, four elements can be identified on a RL system: a policy, a reward function, a value function, and, optionally, a model of the environment.

- Policy: a policy maps a given perceived state of the environment to an action that has to be taken. In some cases the policy may be a simple function or lookup table, whereas in others it may involve extensive computation. The policy is the core of a reinforcement learning agent in the sense that it alone is sufficient to determine behavior. In general, policies may be stochastic.

- Reward function: this element defines the goal in a RL problem. A reward function asigns a state-action pair of the environment to a single number called reward. This reward determines how much desired a state is. A RL agent's main objective is to maximize the total reward it receives. The reward function must necessarily be unalterable by the agent, however it may serve as a basis for altering the policy. For example, if an action selected by the policy is followed by low reward, then the policy may be changed to select some other

action in that situation in the future. In general, reward functions may be stochastic.

- Value function: this element defines what is considered good in the long run. The value of a state is the total amount of reward an agent can expect to accumulate over the future, starting from that state. In an opposite way, rewards determine the immediate, intrinsic desirability of environmental states, but then the value indicate the long-term desirability of states after taking into account the states that are likely to follow, and the rewards available in those states.

- Model: a model mimics the behavior of the agent's environment. A model helps the agent to plan ahead, since given a state and an action, the model might predict the resultant next state and next reward, this helping the agent to decid on a course of action by considering possible future situations before they are actually experienced. Not all the RL methods require a model.

2.3.2 Temporal Difference Learning

As stated on the previous section, a value function is a state-action pair function which estimates how good a particular action will be in a given state. This can be expressed as:

- $V^{\pi}(s)$: the value of a state s under policy π, which is the expected return when starting in s and following π thereafter.

- $Q^{\pi}(s, a)$: the value of taking action a in state s under a policy π, which is the expected return when starting s, taking the action a and thereafter following the policy π.

The problem at this point is how to estimate these value functions for a particular policy. The reason we want to estimate these value functions is so that they can be used to accurately choose an action that will provide the best possible total reward, after being in that given state. Temporal difference (TD) learning methods can be used to estimate these value functions. If the value functions were to be calculated without estimation, the agent would need to wait until the final reward was received before any state-action pair values can be updated. Once the final reward was received, the path taken to reach the final state would need to be traced back and each value updated accordingly. This can be expressed formally as:

$$V(S_t) \leftarrow V(s_t) + \alpha(R_t - V(s_t)) \tag{2.2}$$

Where S_t is the state visited at time t, R_t is the reward after time t and α is the learning rate.

On the other hand, with TD methods, an estimate of the final reward is calculated at each state and the state-action value updated for every step of the way. Expressed formally:

$$V(S_t) \leftarrow V(s_t) + \alpha(r_{t+1} + \gamma V(s_{t+1}) - V(s_t)) \tag{2.3}$$

Where r_{t+1} is the observed reward at time t+1. Temporal difference learning was used on this thesis work to estimate the partial contribution on each decision step to the overall profit.

2.3.3 On-policy and Off-policy Learning

On-policy temporal difference methods learn the value of the policy that is used to make decisions. The value functions are updated using results from executing actions

determined by some policy. These policies are usually soft and non-deterministic. In this thesis, the meaning of soft is that it ensures there is always an element of exploration to the policy. The policy is not so strict that it always chooses the action that gives the most reward. Three common policies are used, ϵ-soft, ϵ-greedy and softmax.

Off-policy methods can learn different policies for behaviour and estimation. Again, the behaviour policy is usually soft so there is sufficient exploration going on. Off-policy algorithms can update the estimated value functions using hypothetical actions, those which have not actually been tried. This is in contrast to on-policy methods which update value functions based strictly on experience. What this means is off-policy algorithms can separate exploration from control, and on-policy algorithms cannot. In other words, an agent trained using an off-policy method may end up learning tactics that it did not necessarily exhibit during the learning phase.

Action Selection Policies

As mentioned above, there are three common policies used for action selection. The aim of these policies is to balance the trade-off between exploitation and exploration, by not always exploiting what has been learnt so far. These common strategies are:

- ϵ-greedy: most of the time the action with the highest estimated reward is chosen, called the greediest action. Every once in a while, with a small probability ϵ, an action is selected at random. The action is selected uniformly, independant of the action-value estimates. This method ensures that if enough trials are done, each action will be tried an infinite number of times, thus ensuring optimal actions are discovered.

- Softmax: one drawback of ϵ-greedy is that it selects random actions uniformly. The worst possible action is just as likely to be selected as the second best.

Softmax remedies this by assigning a rank or weight to each of the actions, according to their action-value estimate. A random action is selected with regards to the weight associated with each action, meaning the worst actions are unlikely to be chosen. This is a good approach to take where the worst actions are very unfavourable.

2.3.4 Q-Learning

Q-Learning, the RL technique used on this thesis work, is an Off-policy algorithm for temporal difference learning. It can be proven that given sufficient training under any -soft policy, the algorithm converges with probability 1 to a close approximation of the action-value function for an arbitrary target policy [Barto, 1998]. Q-Learning learns the optimal policy even when actions are selected according to a more exploratory or even random policy. The procedural form of the algorithm is:

The parameters used in the Q-value update process are:

- Learning rate α, set between 0 and 1. Setting it to 0 means that the Q-values are never updated, hence nothing is learned. Setting a high value such as 0.9 means that learning can occur quickly.
- Discount factor γ, also set between 0 and 1. This models the fact that future rewards are worth less than immediate rewards. Mathematically, the discount factor needs to be set less than 0 for the algorithm to converge.
- Maximum reward that is attainable in the state following the current one. *i.e* the reward for taking the optimal action thereafter.

2.4 POWER TAC: A MULTIAGENT, MULTIMARKET SIMULATOR

Power TAC is a competitive simulation that models a liberalized retail electrical energy market, where competing business entities or brokers offer energy services to customers through tariff contracts, and must then serve those customers by trading in a wholesale market. Brokers are challenged to maximize their profits by buying and selling energy in the wholesale and retail markets, subject to fixed costs and constraints. Costs include fees for publication and withdrawal of tariffs, and distribution fees for transporting energy to their contracted customers. Costs are also incurred whenever there is an imbalance between a brokers total contracted energy supply and demand within a given time slot. The simulation environment models a wholesale market, a regulated distribution utility, and a population of energy customers.

Many countries, mainly on the European Union, have issued directives aiming to gradually open the electricity market for all member states [Farahmand et al., 2012]. Such actions made significant contributions towards the creation of internal electricity markets, which propose an alternative to the monopoly model that most countries around the world use to distribute electric energy to its citizens. An open electricity market offers many benefits to both governments and energy consumers, namely high energy distribution efficiency, price reductions, higher service standards and enhanced competitiviness. On the other hand, more than a dozen states in the United States have introduced retail electricity competition; however there is not enough research done in order to fully understand the impact of a competition on an open energy market in terms of supply, demand, pricing and metering [Joskow and Tirole, 2006].

Across energy markets, small producers will emerge as key players as they

compete to allocate their produced energy into the market. These producers will introduce, by aggregation, large amounts of electricity produced by means of solar panels, wind generators, and some other more classic sources [Lund et al., 2012]. However, a common denominator among renewable sources is a high uncertainty related to the amount of energy that will be produced, and hence, committed. On the other hand, and regardless of the production means, energy consumers require to power up their homes, offices and industries, generating stress over the energy production. And yet, smartgrid metering technologies will allow consumers to shift their energy consumption to a more convenient time of the day in order to achieve a better cost efficiency. All these aspects of an open energy market may cause serious energy imbalance issues on the market, that might lead to a lack of energy on certain places of the network or an overproduction on others, or, on the worst-case scenario, a total failure of the energy network due to greedy strategies or to a lack of proper rules and regulations.

Even if the benefits of a well implemented competition-driven open energy market are bold, the consequences of an incorrect handling of this market may cause serious problems, such as massive blackouts, with the potential to impact a country‘s economy and harm governments, citizens and industries. Therefore, a safe regulation entity is required to issue the set of rules that promote competitiviness among brokers, discouraging unbalance-generating strategies, without compromising the energy requirements of consumers. This regulation entity may be able even to intervene with contention measures, which may include a penalization to the originating brokers, to prevent that a small imbalance issue grows enough to cause a major problem.

The Power TAC platform was developed to study all the latter issues. This platform is a complex simulator that creates an open energy market where consumers require energy, which is provided by energy producers through brokers, regulated by an entity that grants a smooth flow of the transactions on a wholesale market and

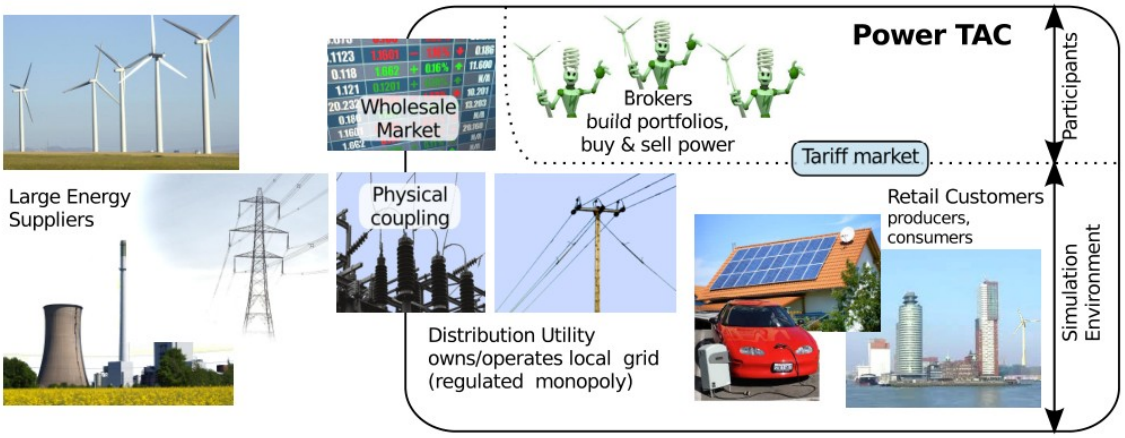

Figure 2.2: General Power TAC view

on a tariff market.

2.4.1 Brokers

The brokers are the main actors of the simulator. Fig. 2.7 provides an overview of which tasks brokers need to accomplish within each timeslot. Typically, this is a three step process: trading in the wholesale market, portfolio development and balancing of energy supply and demand [Babić, 2012] . At the end of each timeslot, or at the end of a given amount of timeslots, the broker can asess its decisions to determine future actions. It is important to mention that the specific order of some activities may not be as rigid as shown in Fig. 2.7.

2.4.2 Tariff Characteristics

A tariff is a contract that allows a business entity or broker to offer energy services to customers [Ketter et al., 2013]. On this scope, customers can be either energy producers or consumers. Having a variety of tariffs contrats offered by many brokers, customers will evalualuate them to decide which one is the most convenient. This convenience is related to the tariff's characteristics. The main aspect of a tariff is

price, an amount paid, by/to the broker, when a MWh [1] of energy is consumed or produced. Other tariff characteristics evaluated by the customers are expiration dates, signup payment, early withdraw payment and even energy origin; a customer might assign a higher profit to a tariff whose origin is the wind or the sun. In the same fashion, customers might feel more attracted to a tariff if it offers them a signup bonus, or feel less attracted for a tariff which requires them a withdraw payment. All the aspects of tariffs available on the PowerTAC simulator are shown on Fig.2.3.

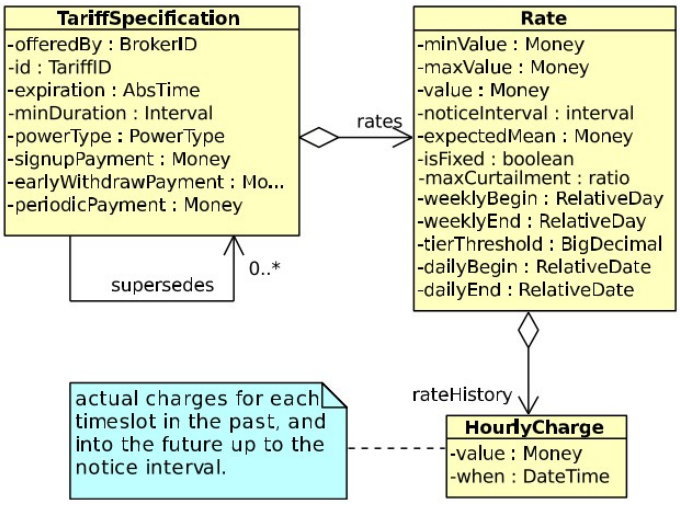

Figure 2.3: Tariff Structure

2.4.3 Tariff Evaluation and Customer Models

In order for potential customers to make comparisons between tariffs and make an informed decision they need some means to evaluate them. For a tariff to be the best choice for any given customer it must provide a utility on some future horizon, where the length of the horizon depends on the customer's preferences. This section describes the process to decide when to consider new tariffs and which tariff to choose given the customer's preference and the tariff's utility.

[1] MWh is the standard energy unit on the wholesale market on Power TAC. However, the standard unit for the tariff market is Kwh. This thesis will use MWh consistently across the whole document.

CUSTOMER'S UTILITY PERCEPTION

The utility perception u_i of a tariff is the parameter that provides a common unit for a customer to determine if a tariff is good or bad. The utility is a function of per-kWh payments $p_{v,i}$, periodic payments $p_{p,i}$, a one time signup payment $p_{signup,i}$, a potential withdrawal payment $p_{withdraw,i}$, which applies in case a customer decides to withdraw its subscription before the tariff's minimum duration, and an inconvenience factor x_i, which accounts for the inconvenience of switching subscriptions:

$$u_i = f(p_{v,i}, p_{p,i}, p_{signup,i}, p_{withdraw,i}, x_i) \tag{2.4}$$

Customers are characterized by their preferences, and therefore by how they evaluate function f, but in general it is the normalized difference between the cost of using the default tariff and the cost of the evaluated tariff, minus the inconvenience factor, as shown on equation 2.9. The default tariff refers to the tariff published by a default broker, which can be thought as the central power company, which sets the market minimum and maximium tariff prices by publishing default tariffs available to all the customers. Therefore, even on a scenario when a single broker publishes one consumption and one production tariff, there are at least four tariffs, two belonging to the default broker and two belonging to the single broker.

The normalized cost difference $n^C_{c,i}$ for consumption tariffs is then the difference between the cost of the default tariff and the proposed or current tariff, normalized by the cost of the default tariff, just as stated on equation2.5:

$$n^C_{c,i} = \frac{cost_{default} - cost_i}{cost_{default}}, \tag{2.5}$$

where $n^C_{c,i}$ represents the normalized cost (c) for the consumption (C) tariff (i) being evaluated.

Since energy consumption is represented by a positive value from the customer's point of view, and payments from the customer to the broker are negative values, the cost values for equation 2.5 are negative.

On the other hand, while evaluating a production tariff, it is better to choose a tariff with a larger payout, so the sign of the cost difference is reversed.

$$n^P_{c,i} = \frac{cost_i - cost_{default}}{cost_{default}} \tag{2.6}$$

On both equations 2.5 and 2.6, the term $cost_i$ is defined as follows:

$$cost_i = \sum_{t=0}^{d_e}(C_{e,t,i} \cdot p_{v,i,t} + p_{p,i}) + (p_{signup,i} + F_d \cdot p_{withdraw,i} + p_{withdraw,0}), \tag{2.7}$$

where:

- $p_{signup,i}$ is a one-time signup payment to tariff i,
- $p_{withdraw,i}$ a potential one-time withdrawal payment from tariff i in case the customer withdraws its subscription before the tariffs minimum duration. If the customer decides to withdraw its suscription after the minimum duration was achieved, then this parameter is zero,
- $p_{withdraw,0}$ is the immediate cost of withdrawing the current tariff 0 to suscribe to tariff i,
- $p_{v,i,t}$ is the per MWh payment at time t. The variable t appears on this term in equation 2.7 but not in equation 2.4 because the latter is not considering a specific time, while the first one does,
- $p_{p,i}$ represents the periodic payments customer have to pay to keep their subscription to tariff i,

- d_e is the expected time that the evaluating customer will require to be suscribed to tariff i,
- $C_{e,t,i}$ is an energy estimate usage over the expected duration $t = [0, ..., d_e]$,
- $F_d = min(1.0, d_i/d_e)$ is a discount factor that adds a premium to shorter commitments intervals d_i.

Equation 2.7 sums up several parameters to account for the customer's cost to suscribe to the new tariff i. The first term considers the expected energy usage $C_{e,t,i}$, multiplying it by the payment per unit of energy consumed $p_{v,i,t}$, and adding the periodic payment $p_{p,i}$ that the customer has to pay at time t. The higher any of these paremeters is, the higher will be the cost that the customer will perceive. The second term considers the cost to suscribe $p_{signup,i}$ to the new tariff i, as well as the possible cost $p_{withdraw,0}$ incurred by withdrawing a previous tariff to suscribe to tariff i. If the minimum subscription time of tariff zero is already completed, then this term will become zero as well. The remaining term $F_d \cdot p_{withdraw,i}$ considers the cost when the customer decides to withdraw tariff i in the future. If the minimum commitment period $d_i = 0$, which means that the customer is not required to keep its subscription to tariff i for any minimum period, then F_d becomes zero as well. If $d_i > d_e$ (meaning that the minimum commitment period required by the tariff is larger that the time the customer is expecting to keep its subscription to tariff i) then $F_d = 1$, and the full value of $p_{withdraw,i}$ will be discounted. For any value for d_i between 0 and d_e, it will hold that $F_d < 1$; thus dampening the value of $p_{withdraw,i}$, and reducing the tariff cost perception for the customer.

Now we will finish to revise equation 2.6. The variable $cost_{default}$ is defined by equation 2.7, but this time replacing with a zero its second term, because default tariffs only include a per-KWh payment and a periodic payment. Equation 2.8 shows this change:

$$cost_{default} = \sum_{t=0}^{d_e} (C_{e,t,default} \cdot p_{v,default,t} + p_{p,default}) \tag{2.8}$$

Where

- $p_{v,default,t}$ is the per MWh payment at time t required by the default tariff,
- $C_{e,t,default}$ is an energy estimate usage over the expected duration $t = [0, ..., d_e]$ and
- $p_{p,default}$ represents the periodic payments customer have to pay to keep their subscription to the default tariff.

Since the tariffs created on this thesis work are priced-based only, the latter equation is used as well by the customers to evaluate the offered tariffs.

Finally, utility is the normalized cost difference less the inconvenience factor:

$$u_i = n_i - w_x \cdot X_i, \tag{2.9}$$

where:

- $w_x \in [0, 1]$ is a static atribute of individual customers, selected from a uniform distribution,
- n_i represents the normalized consumption or production cost difference as stated by equations 2.5 and 2.6, and
- $X_i \in [0, 1]$ is a linear combination of factors that penalize tariff features such as variable pricing and tiered rates. However, it is important to mention that none of these features was used during this thesis work, so for all purposes the latter value can be considered as one.

INERTIA

Customers do not always consider withdrawing to an old tariff and suscribing to a new one, even if the new one provides a better utility. This behaviour is real and is related to the undesired work required to evaluate a new tariff, and is modeled in Power TAC by an inertia factor which determines the probability that a customer will not evaluate tariffs during a tariff-publication event. Inertia is defined as:

$$I_a = (1 - 2^{-n}) \cdot I, \tag{2.10}$$

where:

- I_a is the portion of the population that will not evaluate a new tariff and possibly suscribe to it,
- n is a count of the tariff publication cycles starting at 0 and
- $I \in [0, 1]$ is an inertia constant.

At the first tariff-publication cycle I_a has its maximum value, thus customers will evaluate all tariff offerings, however, as n increases, their interest will be reduced, and they will be less likely to evaluate new tariffs.

RATIONALITY

Customers are not entirely rational, which means that they will not always choose the best tariff. Therefore a smoother decision rule is used by Power TAC, which is based on the multinomial logit choice selection model, which allocates the selection choice proportionally over multiple similar tariffs. The logit choice model assigns probabilities to each tariff, t_i, from the set of evaluated tariffs T as follows:

$$P_i = \frac{e^{\lambda u_i}}{\sum_{t \in T} e^{\lambda u_t}} \tag{2.11}$$

where:

- P_i is the probability of choosing the best tariff i,
- λ is represents the rationality of the customer,
- u_i the utility of the evaluated best tariff i and
- u_t the utility for tariff t

When $\lambda = 0$, the customer is not rational at all, and thus, chooses any of the evaluated tariffs randomly. If $\lambda = \infty$ the customer will always choose the tariff with the highest utility, therefore making more rational choices. The customer models, which will be described on the next section, include groups of members with 3 different values of λ. This feature creates a more realistic simulation, where customers not always choose the best option for any reason. The rationality of the customers depends as well on the amount of tariffs that are published and, most of all, on the utility differences of these tariffs; an example can illustrate this good enough.

Suppose we have two flat consumption tariffs and one customer X which has not yet selected a tariff. This customer can choose between a tariff A and a tariff B, where tariff A has a price of 0.50 units and tariff B one of 0.40 units. The probability of picking the best tariff per λ is shown in Fig.2.4.

Now lets suppose that we have the same customer X, but now the customer can choose between 20 flat tariffs. Tariff A has the highest utility with a price of 0.40, while the remaining 19 have a price of 0.50 units, thus earning to the customer a lower utility. The probability of picking the best tariff as a function of λ for this case is depicted on Fig.2.5 .

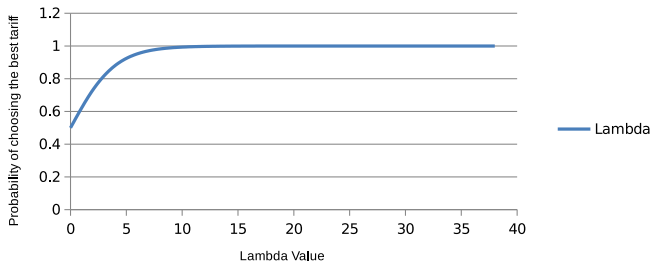

Figure 2.4: Probability of choosing the best tariff per λ when 2 tariffs are available.

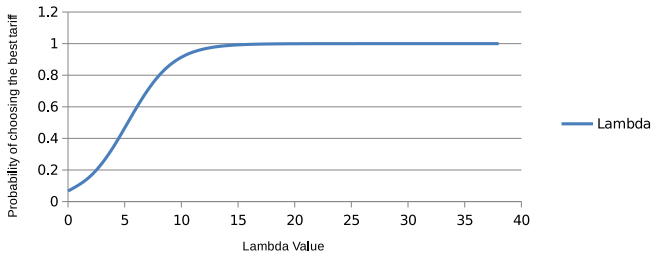

Figure 2.5: Probability of choosing the best tariff per λ when 20 tariffs are available.

It is clear than for 20 tariffs a higher λ is needed to have the same probability of choosing the best tariff, in this case tariff A.

To show that the differences among the customer's computed utility for several tariffs has a heavy impact on the probability of choosing the best tariff, lets consider a third example based on the previous example. Now the price of tariff A is 0.25, so, the difference between tariff A and the remaining is larger than in the previous example. The probability of choosing tariff A for this example is shown in Fig. 2.6.

It is clear that a bigger difference between the tariff's prices results in higher probabilities of picking the best tariff.

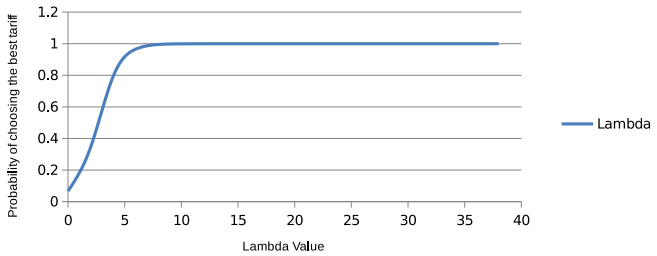

Figure 2.6: Probability of choosing the best tariff per λ when 20 tariffs are available and there is a large differente in tariff prices.

CUSTOMER MODELS

There are a wide variety of customers that require electric energy, and they have different profiles and consumption patterns. These two attributes determine which characteristics might result desirable for them. Customers in PowerTAC are aggregated on Customer Models, and there are three types:

- Household (or Residential) consumers: the most common type of consumer. Every family or person occupying a house is substantially an active household consumer. Even though as single customers have low consumption, when aggregated in big groups their load can be as much as one factory complex. Household customers may as well produce energy for self consumption by using photovoltaic panels, and this energy can be sold as well.

- Business consumers: this consumer category contains all the small or medium businesses, small industries and office complexes. The business consumers demand a greater power supply than residential consumers but not as much as big enterprises and factory complexes.

- Industry consumers: the most energy-consuming customers. Industries use high-voltage power lines and include large manufacturing plants and factories such as chemical plants, computer chip manufacturers or car industries.

2.5 COLD BROKER: A SCALABLE BROKER SCHEME

Trading energy is a complex process, for this reason Power TAC splits it on smaller consecutive processes; this processes are shown on Fig. 2.7. Each process occurs on a timeslot, each timeslot represents one hour of simulated real-time.

- Trading process: where the broker trades on the wholesalemarket to acquire or sell energy with commitments due on a period ranging from 1 to 24 hours. This market was not investigated during this thesis.
- Portfolio development process: this is the most important process for this thesis project, because on this process is precisely where the broker publishes its tariffs and where the customers decide if they evaluate and suscribe to a new tariff or not, as described on Sec. 2.4.3.
- Balancing process: managed by the distribution utility (DU), which represents a regulating corporation. The DU makes the broker accountable for the use of the distribution network depending on the amount of energy traded, this reflects the fact that, on a real market, the broker does not own the distribution network and therefore, has to pay a fee for using it. The DU does as well the job of keeping a healthy balance across the whole network, by charging unbalanced brokers a fee proportional to their inbalance. This mechanism creates an incentive for brokers to keep a balance between produced and consumed energy.
- Accounting service process: is where the broker receives its balance and market position after clearing every single transaction on the current timeslot.

An analog approach to design a broker capable of handling all these processes is to split it as well, and introduce modules, or experts, which handle each process

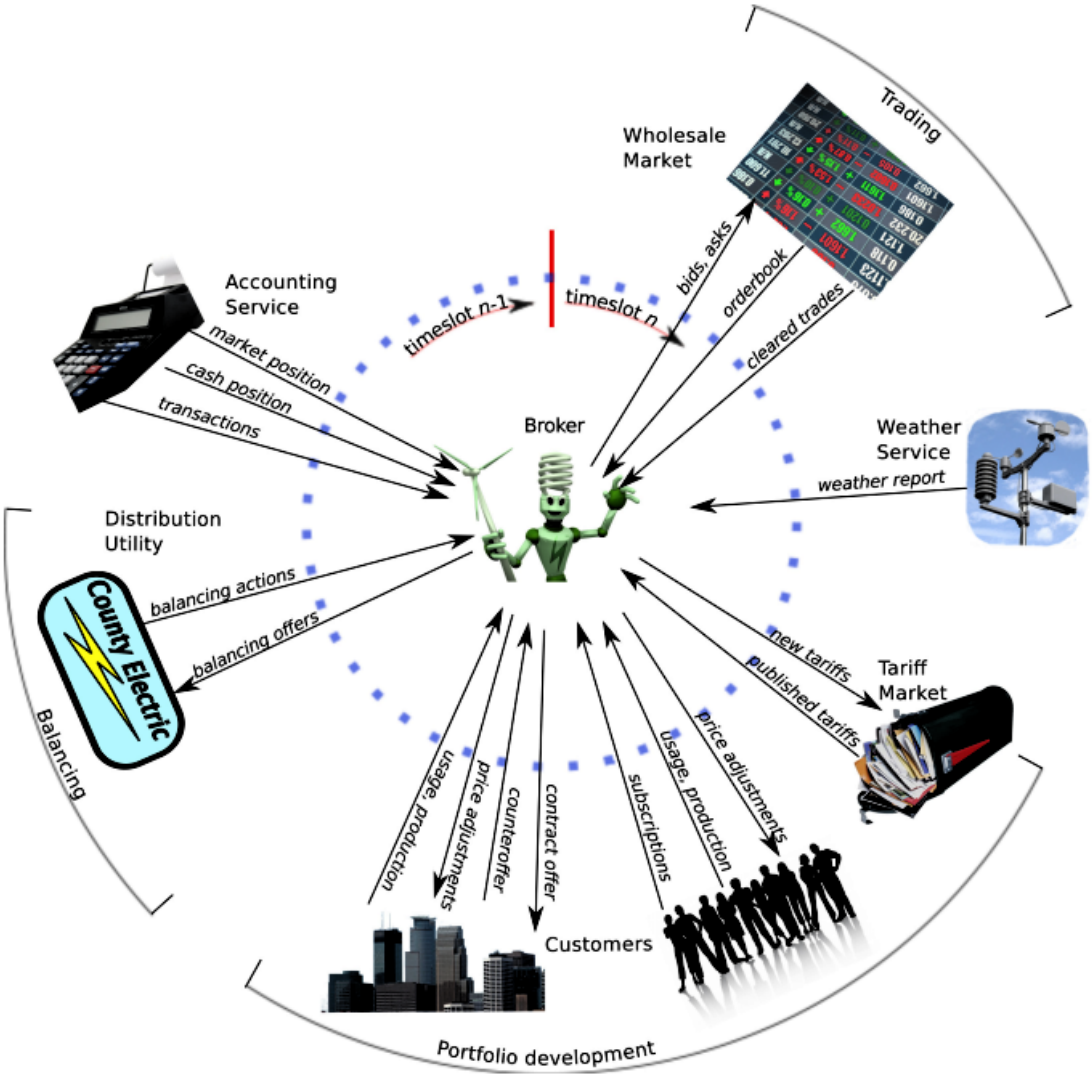

Figure 2.7: Energy trading process on PowerTAC

efficiently. For this reason COLD broker was developed. COLD broker is a scalable scheme which uses experts as bottom-up building blocks to conform an energy broker, where each expert focuses on a single task. The COLD broker scheme is shown on Fig. 2.8. The two main building blocks of the broker are the the market expert and the tariff expert, being the latter the one developed on this thesis project and the one grayed on the figure. The three boxes outside COLD broker represent those models with which COLD broker interacts, and the lines associate a model to an

expert within COLD broker. The wholesale market model interacts with the market expert, the distribution utility model will interact with the balance expert, and the customer models interact with both the market and the tariff expert.

The wholesale market, distribution utility and customer market models are simulated by Power TAC and react to the decisions made by COLD broker, or any other broker in the competition. Each expert is designed using an object-oriented structure, which allows us to develop and test various versions of each broker easily. This structure was used as well to create those brokers against the broker developed on this thesis was tested against. It is important to mention that this structure will serve as a solid base for any related future work.

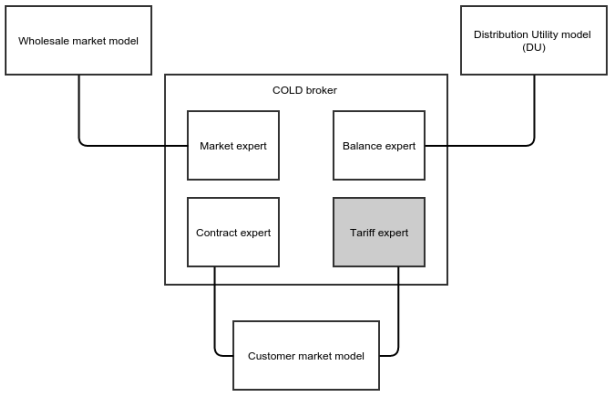

Figure 2.8: Energy trading process on PowerTAC

Chapter 3

Related Work

This chapter is divided into three sections. The first section is an introduction to smart grid to explain its main characteristics but, most of all, why it is important and how it is related to energy markets. The second section reviews the most relevant approaches taken to represent energy markets in order to create tariffs. The last section describes the most common tariff schemes available. Table 3.1 briefly describes the most relevant related work, while table 3.2 shows a point by point comparison with the work proposed by Reddy.

Reference	Brief Description
[NIST, 2012]	States a general description about smartgrid and defines how it fits into a liberalized energy market.
[Farhangi, 2010]	Describes the main differences between smartgrid, as a bidirectional distributed network against, and the current energy network.
[Gordijn and Akkermans, 2007]	Documents evidence on countries where energy markets are liberalized that shows the positive impact that these type of markets generate for producers and consumers.
[North et al., 2002]	Describes a free energy market simulator which focuses on the wholesalemarket.
[Keppo and Räsänen, 1999]	Proposed a pricing tariff scheme which increases a tariff's cost as the forecasted consumption uncertainty for a given customer increases.
[Reddy and Veloso, 2011]	Proposes an energy market model which this thesis work used as the reference to propose a new model. This paper also defines the utility function that this thesis work utilized to measure broker's behavior.

Table 3.1: Most relevant related work.

3.1 Smartgrid and Deregulated Markets

A comprehensive definition proposed by the National Institute of Standards and Technology (NIST) states that a smartgrid is a modernized grid that enables bidirectional flows of energy and uses two-way communication and control capabilities

Feature	Reddy	COLD broker
Simulation Environment	Custom made	Power TAC
Energy Consumption	Fixed	Determined by customer needs
Actions	Non Market-bounded	Market- bounded
Number of states	5	4
Number of actions	6	8
Tariff Price Range	From 0.01 to 0.20 with 0.01 increments.	From 0.015 to 0.5 without a fixed increment
Market Model	MDP	MDP
Learning Technique	Q-Learning	Q-Learning
Learned Attribute	Tariff Price	Tariff Price

Table 3.2: Comparison with Reddy's work

that will lead to an array of new functionalities and applications. Unlike today's grid, which primarily delivers electricity in a one-way flow from generator to outlet, the smart grid will permit the two-way flow of both electricity and information [NIST, 2012]. Table 3.3 shows a comparison between the features of the existing grid and those of a smartgrid.

Existing Grid	Smartgrid
Electromechanical	Digital
One way communication	Two-Way Communication
Centralized Generation	Distributed Generation
Hierarchical	Network
Few sensors	Sensors throughout
Blind	Self-Monitoring
Manual restoration	Self-Healing
Failures and blackouts	Adaptive and Islanding
Manual Check/Test	Remote Check/Test
Limited Control	Pervasive Control
Few customer choices	Many customer choices

Table 3.3: The smartgrid compared with the existing grid

The characteristics of a smartgrid are very different from those on today's energy infrastructure. The existing grid is unidirectional in nature, and due to the hierachical topology of its assets, it suffers from domino-effect failures, that can lead to massive blackouts. It can't prevent in no way service malfunction issues and neither do anything to fix them. On the other hand, the smartgrid will provide to the energy companies a full visibility and pervasive control over their assets and services, it will be able as well to self-heal and make the grid resilient to system anomalies. All these features have the potential to drastically reduce the chances of massive failures. Finally, it will provide the means to empower consumers, producers and energy companies to define and realize new ways of engaging with each other to perform energy transactions across the system, causing it to be sustainable [Farhangi, 2010].

Because of all these reasons, the adoption of green energy sources provided by small producers is being supported by several governments including Portugal,

Netherlands and Germany. [Ringel, 2006] proposed a first model to guarantee that electricity generated with the use of renewable energies can be completely feeded into the power network. On this model, the lawmaker obliges regional or national transmission system operators (TSO) to feed-in the full production of green electricity at politically fixed prices differing according to the various generation sources (wind, hydro, etc.). These tariffs cover the cost disadvanatage of the renewable energy sources and are calculated so as to grant an investment bonus to the green power producers.

There is documented evidence that a smartgrid infrastructure, along with the efforts to promote investment on green energy sources by using convenient tariff schemes, could provide good results in terms of flexibility for customers and contamination reduction [Gordijn and Akkermans, 2007]. Along this line Gordijn analyzed several case studies in european markets, and one of them was Spain, where the electricity market is fully liberalized since 2003, smartgrid‘s load shifting feature could reduce the electricity bill of a final customer by 15%.

Norway is another interesting case. This country, as many others around the world, is experiencing a steady energy consumption growth. The energy generation on this country depends mainly on hydropower; even though this has historically contributed to low electricity prices, this makes the country very vulnerable to dry years, which are becoming increasingly common due to climate disorders. On the other hand, there is limited space to build new large-scaled hydropower plants, therefore alternative options are required. The study made by Gordijn proposes a total deregulated market environment where customers own generating facilities and produce electricity for their own consumption, while the surplus electricity production output is sold on the power market. The central actors in this model are the local producers, who generate electricity that is sold to the electricity supplier, using distribution and metering services provided by the local distribution system operator. The producers receive payment from the supplier for the electricity output, accord-

ing to the metered data. The producers pay a network tariff for feeding electricity into the distribution system and also for distribution and metering services. This scheme is different from the model used on this thesis, where it is the broker who is accountable for such payments.

3.2 Energy Markets and Tariff Generation

Some research has been done regarding energy markets and tariff generation, most of it has taken the form of simulation platforms. North et al. created an agent-based simulation of an electric power system including the consumers and the producers [North et al., 2002]. The agents' objectives were characterized by a utility function and a specific complex adaptive system approach was used to represent the agent's learning capabilities. The electronic market was focused on the wholesale market and the bilateral contracts between energy suppliers and energy consumers and not on the tariff market. The study proved that agent-based simulations, where each agent has its own objectives and decision rules, make it possible to represent power markets more accurately than those simulators which rely on an implicit decision maker.

Maenhoudt further analyzed the importance of studying electric power markets, but moreover, the benefits to model the problem as a multi-agent problem [Maenhoudt and Deconinck, 2010]. This study concluded that agent-based simulations are a favourable tool for decision makers in the electric power market, because it comprises not only the economic, but also the social and environmental factors operating in the system.

Keppo and Rasanen created a model to price tariffs in competitive electricity supply markets [Keppo and Räsänen, 1999]. This model uses the value of the customer's electricity consumption pattern. A customer might, for instance, be very consistent on its consumption, so it will have a static consumption pattern. On the

other hand, another customer might have a very variable electricity consumption. Considering this, Keppo's model is based on the future price of the value of this pattern, and showed that customers with a high uncertainty in their consumption pattern should be charged more than less uncertain customers. Their analysis also indicated that deterministic load profiles cannot be applied in a competitive market, because they do not include the consumption risks. Keppo's paper was focused on creating a specific pricing model for a customer's consumption pattern; in contrast, this thesis work focused on creating a strategy that can alter tariff prices to attract customers that are subscribed to competitors; also, this thesis results use experimental evidence via a simulator, while the work presented by Keppo presents results via a mathematical model.

3.2.1 Adaptive Tariff Generation on Competitive Markets

[Reddy and Veloso, 2011] used a simulation approach to investigate a heavily simplified competitive tariff market, where the amount of energy consumed and produced by customers was discretized on blocks, and the daily consumption was a fixed parameter that remained the same through the entire simulation. The paper used agents with 5 different strategies, each of them using different actions to alter tariff prices. The learning strategy learned a Markov Decision Process policy by using Q-learning. The states of the Q-learning algorithm consisted of two heuristic elements. One of them captured the broker's energy balance, determining if more energy was bought than sold or it was the other way around. The second element captured the state of the market by comparing the minimum consumption price and the maximum production price. The paper demonstrated that agents that used the learning strategy overperformed those that used a fixed strategy in terms of overall profit. The author showed this results on several figures plotting utility values. However, statistical parameters were not used to confirm this. This thesis has similarities with this

paper, however, a major difference is that customers in this thesis are more complex and consumption and production patterns are more realistic, since these parameters were not manipulated at all. Also, our broker has more meaningful actions and an enhanced market representation that allows it to react to other brokers actions.

Given that the purpose of this thesis was to develop tariff pricing strategies, the actions only performed changes over tariff rates, however, the model can be used to generate complex tariffs, such as those with bonus signup payments or time-independentt tariffs. Even as these tariff characteristics were not used, they are briefly explained on Sec. 2.4.2

Chapter 4

Tariff-Broker Design

4.1 Tariff-Broker's Problem Statement

The objective of the broker developed on this thesis is to learn a policy which chooses an action at each decision step that supplies a good long term utility, *i.e.* take good actions at every decision step, so as to maximize the sum of immediate utilities gathered over every time step. It is important to mention that the learned policy may not be the optimal, because the simulation environment is not stationary [1], but it will yield COLD broker with a larger utility, compared to that obtained by the competitors. It is important to mention again that COLD broker and its competitors only offer fixed-priced tariffs.

To matematically describe the broker's objective it is necessary to define the following.

On any given competition α there are $T = \{1, 2, ..., N\}$ decision steps and $B = \{B_1, B_2, ..., B_k, ..., B_M\}$ brokers. At the end of α the utility of B_k is defined as:

[1] The simulation environment is not stationary because the same action on the same state will not always yield the same reward. This occurs because of factors such as rationality and inertia, but also because there are customer behaviours modeled with random parameters.

$$U_{B_k} = \sum_{t \in T} (u_t^{B_k}), \tag{4.1}$$

where $u_t^{B_k}$ is the per-timeslot utility for B_k as defined in equation 4.2

So the problem every broker faces is to develop a policy π_k^{max} that maximizes its utility as defined by equation 4.1 to earn a larger profit compared to other brokers. On a competition, the profit obtained by any broker depends on the customers and on the competing brokers, because the first tend to choose the best possible tariff which accomodates their preferences and the latter try to match customers preferences. These aspects create a complex moving target for every broker.

For this reason, the first task to be done in this thesis is to design a broker capable to extract the main market features that give quality information for decision making while accounting for the non-stationarity induced by the competing brokers. The next section explains how the tariff expert broker represents the market and which are the available actions at any given decission step.

4.2 Environment Representation

At the end of each evaluation period any broker, including the learning broker, broker B_k publishes a consumption and a production tariff with prices $P_{t,C}^{B_k}$ and $P_{t,P}^{B_k}$ respectively, where customers can suscribe to. At the end the evaluation period, $\Psi_{t,C}$ and $\Psi_{t,P}$ represent the amount of energy sold or acquired by the broker respectively.

For each evaluation period, the utility function is the one shown on equation 4.2. The first term represents the income total proceedings due to electric energy sale, the second terms corresponds to the amount paid to producers, and the third term represents an inbalance fee.

$$u_t^{B_k} = P_{t,C}^{B_k}\Psi_{t,C} - P_{t,P}^{B_k}\Psi_{t,P} - \theta_t|\Psi_{t,C} - \Psi_{t,P}| \tag{4.2}$$

Each term in equation 4.2 represents either a monetary income or outcome. So the whole utility represents money, and its currency is dollars. All three terms multiply a price per energy unit by an energy amount, yielding a monetary unit. If the difference $\Psi_{t,C} - \Psi_{t,P}$ equals zero, then the broker sold exactly the same amount of energy it bought, so the energy inbalance is zero; and for this reason the inbalance fee is zero as well, disregarding the value of θ_t. The variable θ_t is the amount the broker has to pay to the DU per each unit of energy inbalance it generated on the evaluation period.

The broker has full control of two parameters of equation 4.2: $P_{t,C}^{B_k}$ and $P_{t,P}^{B_k}$. These are the consumption and production prices respectively. The rest of the terms are either determined by the distribution utility or depend on customers subscription decisions. To set the consumption and production prices, the first step was to determine some key elements belonging to the tariffs published by other brokers; namely: maximium and minimum consumption prices, and maximum and minimum production prices. These are:

Minimum consumption price:

$$P_{t,C}^{min} = min_{B_k \in B \setminus \{B_L\}} P_{t,C}^{B_k} \tag{4.3}$$

Maximum consumption price:

$$P_{t,C}^{max} = max_{B_k \in B \setminus \{B_L\}} P_{t,C}^{B_k} \tag{4.4}$$

Minimum production price:

$$P_{t,P}^{min} = min_{B_k \in B \setminus \{B_L\}} P_{t,P}^{B_k}, \tag{4.5}$$

Maximum production price:

$$P_{t,P}^{max} = max_{B_k \in B \setminus \{B_L\}} P_{t,P}^{B_k}, \tag{4.6}$$

Equations 4.3, 4.4, 4.5 and 4.6, B_L are valid for any given broker B_k, but to specifically refer to the values observed by the learning broker, B_k will be replaced by B_L. In this way, the minimum and maximum prices include the list conformed by the prices of all the other brokers but not the prices of the learning broker B_L.

Now we can formulate the proposed MDP which use the equations just defined. As stated before B_L is the learning broker for which we develop an action policy using the an MDP based framework and reinforcement learning. The MDP for B_L is defined as:

$$M^{B_L} = \langle S, A, P, R \rangle \tag{4.7}$$

where:

- $S = \{s_i : i = 1, ..., I\}$ is a set of states,
- $A = \{a_j : j = 1, ..., J\}$ is a set of actions,
- $P(s, a) \rightarrow s\prime$ is a transition function and
- $R(s, a)$ equals $u_t^{B_{k=L}}$ and represents a reward function.

Then $\pi : S \rightarrow A$ defines an MDP action policy.

4.2.1 STATES

The proposed state space S is the set defined by the following tuple:

$$S = \langle PRS_t, PS_t, CPS_t, PPS_t \rangle \tag{4.8}$$

where:

- $PRS_t = \{rational, inverted\}$ is the Price Range Status at time t,
- $PS_t = \{shortsupply, balanced, oversupply\}$ is the Portfolio Status at time t,
- $CPS_t = \{out, near, far, veryfar\}$ is the Consumers Price Status and
- $PPS_t = \{out, near, far, veryfar\}$ is the Producers Price Status,

DESCRIPTION OF TUPLE PARAMETERS PRS_t AND PS_t

The values PRS_t and PS_t capture the relationship between the maximum production price and the minimum consumption price, and the balance of the broker B_L; respectively, and are defined as follows:

$$PRS_t = \begin{cases} rational & \text{if } P_{t,C}^{min} > P_{t,P}^{max} \\ inverted & \text{if } P_{t,C}^{min} \leq P_{t,P}^{max} \end{cases} \tag{4.9}$$

$$PS_t = \begin{cases} balanced & \text{if } \Psi_{t,C} = \Psi_{t,P} \\ shortsupply & \text{if } \Psi_{t,C} > \Psi_{t,P} \\ oversupply & \text{if } \Psi_{t,C} < \Psi_{t,P} \end{cases} \tag{4.10}$$

One should remember that the maximum production price and the minimum consumption price, from B_L point of view, are defined considering the list of all the available production and consumption prices, respectively, of all the other brokers, excluding the prices offered by B_L.

DESCRIPTION OF TUPLE PARAMETERS CPS_t AND PPS_t

These two elements of S encode the price actions of the broker related to the actions of the other brokers. Both CPS_t and PPS_t can take any of these values: *out,close,far,very far* and are defined as follows.

Function definition for CPS_t:

$$CPS_t = \begin{cases} out & \text{if } Top_{ref} \leq P^{B_L}_{t-1,C} \\ near & \text{if } Thres_{ref} < P^{B_L}_{t-1,C} \leq Top_{ref} \\ far & \text{if } Middle_{ref} < P^{B_L}_{t-1,C} \leq Thres_{ref} \\ veryfar & \text{if } P^{B_L}_{t-1,C} \leq Middle_{ref} \end{cases} \tag{4.11}$$

where:

- $Top_{ref} = P^{min}_{t,C}$,
- $Middle_{ref} = \dfrac{P^{min}_{t,C} + P^{min}_{t,P}}{2}$
- $Thres_{ref} = \dfrac{Top_{ref} + Thres_{ref}}{2}$

The purpose to define reference variables Top_{ref}, $Middle_{ref}$ and $Thres_{ref}$ was to discretize the continuous consumption price values available in the range [0.015, 0.5]. One should remember that this is the price range where brokers can set their prices. Now lets define PPS_t.

Function definition for PPS_t:

$$PPS_t = \begin{cases} out & \text{if } Bottom_{ref} \geq P^{B_L}_{t-1,P} \\ near & \text{if } Thres_{ref} \geq P^{B_L}_{t-1,P} > Bottom_{ref} \\ far & \text{if } Middle_{ref} \geq P^{B_L}_{t-1,P} > Thres_{ref} \\ veryfar & \text{if } P^{B_L}_{t-1,P} \geq Middle_{ref} \end{cases} \tag{4.12}$$

where:

- $Bottom_{ref} = P_{t,P}^{min}$,
- $Middle_{ref} = \frac{P_{t,C}^{min} + P_{t,P}^{min}}{2}$
- $Thres_{ref} = \frac{Bottom_{ref} + Thres_{ref}}{2}$

Similarly as described for CPS_t, the references $Bottom_{ref}$, $Middle_{ref}$ and $Thres_{ref}$ discretize the values that the production prices can take.

4.2.2 ACTIONS

The set of actions is defined as:

$$A = \{maintain, lower, raise, inline, revert, minmax, wide, bottom\} \tag{4.13}$$

where each of these actions define how the learning agent B_L determines the prices $P_{t+1,C}^{B_L}$ and $P_{t+1,P}^{B_L}$ for the next timeslot t+1. These are the specific details of each action:

- *maintain* publishes the same price as in timeslot t-1 aiming to keep the current MDP state.
- *lower* decreases both consumer and producer prices by a fixed amount ϵ with the purpose of gaining new consumption customers and eliminating some production customers.
- *raise* increases both the consumer and producer prices by a fixed amount ϵ. Has the opposite effect that *lower*.

- *inline* sets the consumption and production prices as $P^{B_L}_{t+1,C} = \lceil m_p + \frac{\mu}{2} \rceil$ and $P^{B_L}_{t+1,P} = \lfloor m_p - \frac{\mu}{2} \rfloor$. The Inline action is market-bounded. Similar to bottom but more aggressive, because it sets a very cheap price to buy energy from producers and also a very attractive low price for consumers.

- *revert* moves the consumption and production prices towards the midpoint $m_p = \lfloor \frac{1}{2}(P^{min}_{t,C} + P^{min}_{t,P}) \rfloor$. This is an emergency action which has the objective to quickly gaining consumption and production customers.

- *minmax* sets the consumption and production prices as $P^{B_L}_{t+1,C} = D_{coeff} P^{max}_{t,C}$ and $P^{B_L}_{t+1,P} = P^{min}_{t,P}$, where D_{coeff} is a number on the interval [0.70,1.00] which damps the effect of the minmax action over the consumption price.

- *wide* increases the comsumption price by a fixed amount ϵ and decreases the production price by a fixed amount ϵ.

- *bottom* sets the consumption price as $P^{B_L}_{t+1,C} = P^{min}_{t,C} \dot{Margin}$, where $Margin = 0.90$; and the production price $P^{B_L}_{t+1,P} = P^{min}_{t,P}$. The Bottom action is market-bounded. Similar to inline but less aggressive. Its purpose is to set a consumption price just below the minimum consumption price offered by any other broker.

4.2.3 State/Action Flow Example

To illustrate an action's effect over the consumption and production prices, Fig. 4.1 shows a simple simulated flow on a series of actions. The actions appear above the graph. On this hand-made simple scenario COLD broker competes against two brokers, who publish a consumption and production tariff each. The horizontal axis represents the time measured in decision steps, the vertical axis corresponds to the energy price. The dashed lines are fixed references, while the continuous lines are the published prices as described below:

- maxCons: corresponds to $P_{t,C}^{max}$ and is equal to 0.5. It can be assumed that competing broker A published a consumption tariff with this price.
- minCons: corresponds to $P_{t,C}^{min}$ and is equal to 0.4. It can be assumed that competing broker B published a consumption tariff with this price.
- minProd: corresponds to $P_{t,P}^{min}$ and $P_{t,P}^{max}$; which means that the maximum and minimum production prices are the same and is equal to 0.015. It can be assumed that both brokers A and B published a production tariff with this price.
- Cons: corresponds to the consumption price published by COLD broker.
- Prod: corresponds to the production price published by COLD broker.

COLD broker will bound the price range of its tariffs in the range $[P_{t,P}^{max}, P_{t,C}^{min}]$. For this reason, none of the actions will lead to a price position outside this range.

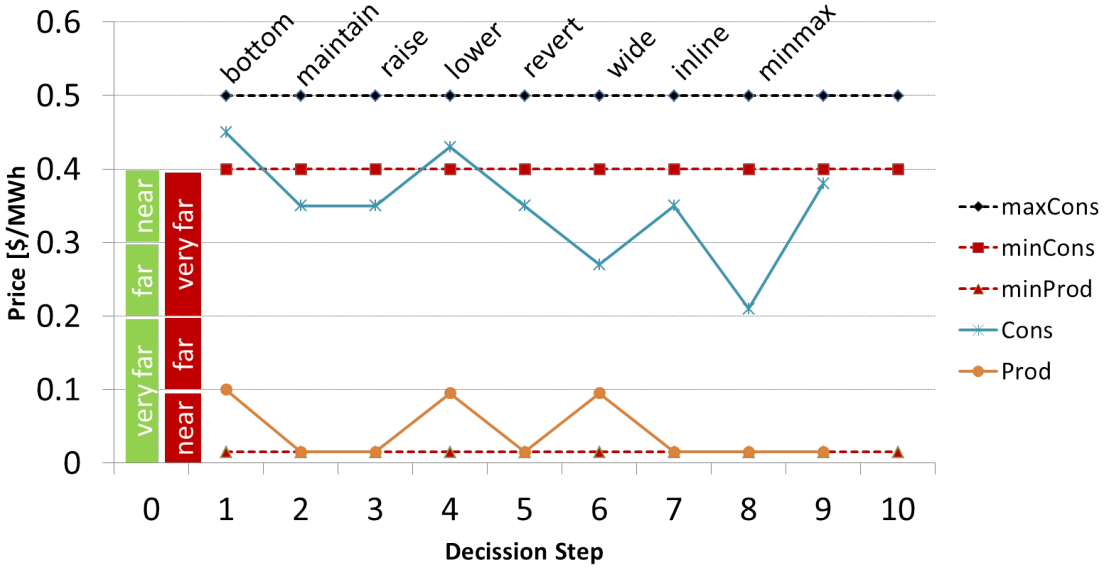

Figure 4.1: A graphical representation of Coldbroker's actions and states.

The continuous lines on Fig. 4.1 are the consumption and production prices published by COLD broker at the given decision step. There are as well some colored blocks at the left side of the graph. The green blocks represent the price region for which CPS_t take a certain value. The red blocks represent the price region for which PPS_t take its own values. Any value outside the boundaries defined by the green and red colored blocks, including the outer edges, is considered as *out*. It is important to mention that this is just an illustrative example, with the only purpose of showing what is the effect of each action on the consumption and production prices, and also to show how these actions change the broker's state.

So, at decision step (DS) 1 COLD broker takes action *bottom*, lowering both consumption and production prices. The effect of the action taken in DS 1 is reflected on DS 2. So, at DS 2 the consumption price is between a price of 0.4 and 0.3; this means that the corresponding values for CPS_t and PPS_t are *near* and *out* respectively. It is important to remember that the action *Bottom* sets the production price to $P^{min}_{t,C}$, which, by definition, corresponds to a value of *out* for PPS_t. At DS 2 COLD broker maintains both prices. At DS 3 the action is *increase*, and both prices are raised. The opposite occurs on DS 4. On DS 5 and DS 6 the actions revert and wide are taken. These are opposite actions. While revert moves the prices towards the center, wide moves them away. Revert sets CPS_t value to *far*. At DS 7 the action inline is executed. This is a more aggresive action compared to bottom, because it sets the consumption price far away from the competitions lowest consumption price; and the production price at the minimum available. It is important to mention that this sequence of actions was hand-made with the only purpose of illustrating the effects of the executed actions.

This simple scenario assumes that the competing brokers are not changing the market state because they do not publish new tariffs. However, when they do they might alter $P^{min}_{t,C}$ or $P^{min}_{t,P}$. This would lead COLD broker to recalculate its market range and therefore publish new tariff prices accordingly, which is what actually

happens.

4.3 Learning Strategy

The learning algorithm used was the Watkins-Dayan [[Watkins and Dayan, 1992]] Q-Learning update rule:

$$\hat{Q}_t(s,a) \leftarrow (1-\alpha_t)\hat{Q}_{t-1}(s,a) + \alpha_t \left[r_t + \gamma \hat{Q}_{t-1}(s', a') \right], \qquad (4.14)$$
$$\max_{\text{a'}}$$

As it will be described later, COLD broker was trained offline in order to obtain an initial Q-Table prior to running the experiments.

4.4 Implementation Challenges

Before stating the specific challenges faced to develop a tariff expert, it is important to mention that this thesis was developed as a component of a more complex and scalable broker, called COLD broker. This broker comprises two main components so far: the market expert and the tariff expert. The experiments executed on this thesis work were done in isolation, using only the tariff expert in order to comprehensively understand the effects of the proposed actions and strategies without any other influence. However, the design and programming was done so that this tariff expert could fit easily on the COLD broker scheme. The COLD broker was tested as whole including the tariff and market experts during the 2014 Power TAC tournament. This stated some additional challenges that had to be worked out. The most important challenge might be the communication between these two experts. To address this issue, the broker's architecture allowed both experts to keep track of and/or update the customer, tariff, price and transaction records online. This modular architecture allows to develop and test the broker easily, so that updates

and improvements in the future does not compromise the whole broker. Now, the specific challenges related to developing a tariff expert are related to several facts. First of all, to represent the state of an energy market at any given moment requires to create or select some features that provide enough relevant information. This task requires a discretization process of several variables, whereas the consumption and production prices are the most important. Tariff prices are not easy to represent because in a real energy market they are continuous over all the real domain. If we were to make a gross discretization, for instance, of 0.05 units over the available price range of 0.015 to 0.5, we would have around 9 different possible values. If we considered 5 brokers, which is far off a real world situation, there would be 9^5 possible states. This is not a viable input for an MDP. The first challenge was therefore to determine which market attributes were appropriate to represent it's dynamics for a good decision. The market state variable PRS_t proposed by Reddy provided a good approach, and this thesis work enhanced this market representation by introducing the variables CPS_t and PPS_t; which provided the broker with a better insight to make decisions. These two market state variables required a full book keeping of others brokers actions, specifically their price attribute.

Several other challenges are directly related to the realistic energy market representation provided by Power TAC. As described on Sec. 2.4.3, the customer models simulated are fairly complex, and include aggregated customers that create populations. Each of these customer models may have a different set of preferences, that include change resillience (inertia parameter I) and the desire of individual customers to choose cheaper or more expesive tariffs (λ parameter). The learning process had to deal with the different preferences among customers, to create prices for tariffs that fit best to all of them.

Another challenge was to analyze the logs and the information generated by the simulation process. Each simulation generates large logs information on almost every aspect of the game, including tariff related events, such as:

- Publication of a new tariff.
- Revocation of a new tariff.
- Customer subscription to a tariff.
- Customer transaction over a given tariff: buy or sell energy.

Each of these aspects may include a monetary value and a traded energy amount, along with an associated customer, an associated tariff and an associated timeslot. The broker accounts for all this information while running on a Power TAC simulation, but it was required to develop tools related to big data analysis to select all relevant information online and process it into tables and plots which supplied enough information for a comprehensive analysis to take place.

The proposed broker's scheme represented a challenge as well. The structure was designed to be scalable and reused, and therefore it required time to be developed and tested for the first time.

The complexity of the simulation platform represented another challenge, mainly because the MDP required a discrete set of attributes. This discretization process was not easy, because there are more than two brokers and there are several variables with continuous values. In this environment a naive discretization might lead to a model which grows exponentially. For instance, if we tried to discretize the continuous values of consumption prices using a non-coarse discretization, the MDP's number of states might grow exponentially as the number of tariffs or brokers increased. Let us remember that states are represented as rows in a Q-Table, while actions are represented as columns. Having this in mind, and considering that when the broker chooses an action a new set of consumption and production prices is proposed, one could tell that if the actions where to choose fixed price values; the number of actions could as well grow exponentially, since prices are continuous values. However the discretization proposed in this thesis to represent the market keeps

its size constant, regardless the number of competing brokers or published tariffs.

Finally, simulation time by itself was another challenge. In order to test the effects of any assumption or algorithm, several simulations had to be done to run statistics. A short simulation required at least one hour, while the longest ones required more than five hours. For this reason, the simulation process was very time-consuming.

Chapter 5

Experimental Results

This section will describe the results obtained by using the market representation and the actions described in Chapter 4. More than 500 experiments were tested only during the experimental phase on this MSc work and this section provides the final and most relevant results. On each experiment a different set of brokers participated. The different brokers are described in Table 5.1.

Broker Name	Description
COLD broker	The learning broker developed on this thesis work.
ReddyLearning	The learning broker proposed by Reddy.
Fixed	Publishes initial tariffs and never updates them again.
Balanced	A fixed-strategy broker which uses the Balanced strategy.
Greedy	A fixed-strategy broker which uses the Greedy strategy.
Random	A broker that uses the same market representation and chooses. random actions over the set available for COLD broker.

Table 5.1: Competing brokers general description

The final experiments included a test where COLD broker competed against the rest of the brokers described on Table 5.1, a test where COLD broker competed against ReddyLearning and another one where COLD broker competed against another version of itself.

Before describing the experiments and their results, the next brief section will describe a series of experiments that were executed to analyze the impact of changing the customer's rationality parameter λ while keeping fixed a pair of consumption and production prices. Even though this analysis was beyond the stated reach of this thesis, it was considered relevant to analyze because it would help to better understand the MDP learning process.

5.1 Analysis on the Rationality of Customers (Lambda Parameter)

As mentioned in Section 2.4.3, the customer's rationality has an impact over the broker's ability to obtain higher utilities. Even as if this is out of this thesis scope, some experiments where done to test the impact of changing the λ lambda parameter. These experiments details are shown on Appendix A . The experiments showed that as the rationality of a customer increases, the tariffs with a slight difference from the reference[1] will be perceived as with high value. Therefore, when a high value of lambda exists, the broker is able to maximize its profit by publishing prices very close to the reference values, *i.e* very expensive consumption prices and very cheap production prices. As the rationality decreases, a larger difference will be required between any given broker's price and the reference price for the customers to perceive it with a high value, or low cost, hence reducing the maximum profit that any broker could obtain.

5.2 Experimental Setup

Now the details of each experiment will be described, starting with the setup and then following with some useful conventions. Besides including or excluding one or

[1] Customers always evaluate tariffs comparing them against another reference tariff

more competing brokers, each experiment's settings was kept the same. Prior to the experiments, both COLD broker and ReddyLearning were trained so as to test them with the best actions that these brokers could learn. Then they were tested against the other brokers.

Training Sessions

COLD broker and ReddyLearning were trained against a Fixed broker for 2,000 timeslots and against the Random broker for 8,000 timeslots. During the training sessions the brokers were adjusted to explore at every decision step, updating their Q table with the obtained reward. The learned Q table for each of the trained brokers was stored and then transfered to the learning brokers to be used on the test sessions.

Test Sessions

The trained Q table was stored and transferred to the brokers to be used on the experiments. The experimental general setup is described below:

- Game length: 3000 timeslots, corresponding to 125 simulated days and 75 decision steps. The first timeslots does not appear on simulations because the simulations start around timeslot 500.
- Tariff publication interval: all brokers publish new tariffs every 50 ± 5 timeslots.
- Tariff types: every tariff publication event, brokers publish one consumption and one production tariff.
- Explore ratio: since the learning process already took place, the brokers followed their learned policy, thus they did not explore during the Test Sessions.

- Tariff price range: tariffs prices are within the interval [0.015,0.5].

Since the target of any broker was to publish attractive tariffs by changing their prices, these were the only attributes that were changed by the brokers during the simulations. The other parameters just mentioned remained constant.

5.3 Conventions

In order to keep easy-readable and reduced tables, some abbreviations were used to designate the names of the values each state can take. The abbreviation consisted on using the first two letters of the value's name, as stated on Table B.1. So, for instance, state representation RaShFaOu stands for state $S = \langle rational, shortsupply, far, out \rangle$. The capital letters were used just to differentiate each state's name easily.

State Attribute	Possible Values (abbreviation)
PRS_t	rational(Ra), inverted(In)
PS_t	shortsupply(Sh), balanced(Ba), oversupply(Ov),
CPS_t	very far(Ve), far(Fa), near(Ne), out(Ou)
PPS_t	very far (Ve), far(Fa), near(ne), out(ou)

Table 5.2: States values and abbreviations

5.4 Book Keeping

In order to analyze the broker's behavior, each one of them generates a detailed log which contains information about the market state, performed actions, tariff prices and yielded utility for each decision step. This information is stored in a file called States. This file stores the data on various columns. Each register includes information about the timeslot, the published consumption and production prices,

the action performed, the MDP state and the reward obtained. The detail on how this information is stored can be found on Appendix B.

5.5 Experiments Description

On this section each of the experiments will be shown. For each experiment a description and an analysis will be provided. The experiments were designed to test COLD broker against specific sets of the competing brokers and itself. The experiments that will be explained are as follows:

- COLD broker vs. All: our learning broker vs. Random, Balanced, Greedy and the learning broker proposed by Reddy, named as ReddyLearning.
- COLD broker vs. ReddyLearning: our learning broker vs. the learning broker proposed by Reddy.
- COLD broker vs. COLD broker: our learning broker vs. another instance of itself.

5.5.1 COLD broker vs. All

This series of experiments included all the brokers. Table 5.3 shows on the first column[2] the market state, as described on section 4.2.1. The next columns show the average utility and its corresponding standard deviation for each of the brokers. The last row shows the overall average and standard deviation for each broker. Fig. 5.1 plots this last row of data. Fig. 5.2 shows the utility for each broker at each timeslot.

[2]This column shows the first two letters of each atribute's name.

	Balanced		Cold		Greedy		ReddyLearning		Random	
State	**Average**	**Std. Dev**	**Average**	**Std. Dev**	**Average**	**Std. Dev**	**Average**	**Std. Dev**	**Average**	**Std. Dev**
RashFaOu			174,153	11,625					102,456	73,811
RashVeOu			173,011	18,419					130,540	
RashFaNe			170,911	20,262					72,790	114,024
RashNeOu			155,069	24,060					99,016	88,380
RashOuFa	261		148,156	16,380	- 7,513	-	147,667		- 6,428	4,495
RashNeNe			132,635	9,086						
RashOuOu			129,775	46,235	- 7,960	125			24,379	48,751
RashOuVe	- 6,313	2,311	126,248				23,663	37,111	10,078	62,824
InshNeNe			122,193	46,484					82,709	41,243
InshOuFa	- 5,511	4,473	111,042	21,857	- 7,786	308			30,420	48,639
InshOuVe	- 11,875	7,083	99,340	40,064	- 7,728	343	1,141	15,419	43,969	45,035
InshNeOu			98,531	41,451					79,228	45,536
InshFaNe			95,388	27,956					92,427	26,106
InshVeNe			91,703	-					106,631	7,481
InshOuNe	- 5,172	2,305	89,993	3,679	- 7,797	288			23,464	41,034
InshFaOu			86,998	53,118					90,494	39,285
InshVeOu			86,447	45,530					112,301	49,444
InshOuOu	- 4,757	3,428	56,708	28,661	- 7,809	258			19,789	35,326
RashFaVe	58,018		52,400		44,510		51,925		80,048	
RaovOuVe	- 15,966	18,831					6,454	17,590		
InovOuOu	- 2,213	8,171								
InshNeVe							145,151			
InovOuVe	- 8,364	4,235					1,762	15,281		
InovOuFa	- 4,015	4,253								
RashNeVe							161,590	53,707		
InshNeFa									56,759	56,130
RashOuNe									2,636	22,328
Summary	**- 7,008**	**8,317**	**107,238**	**54,109**	**- 7,552**	**3,491**	**11,459**	**34,963**	**49,507**	**55,078**

Table 5.3: Average and standard deviation per state for each broker

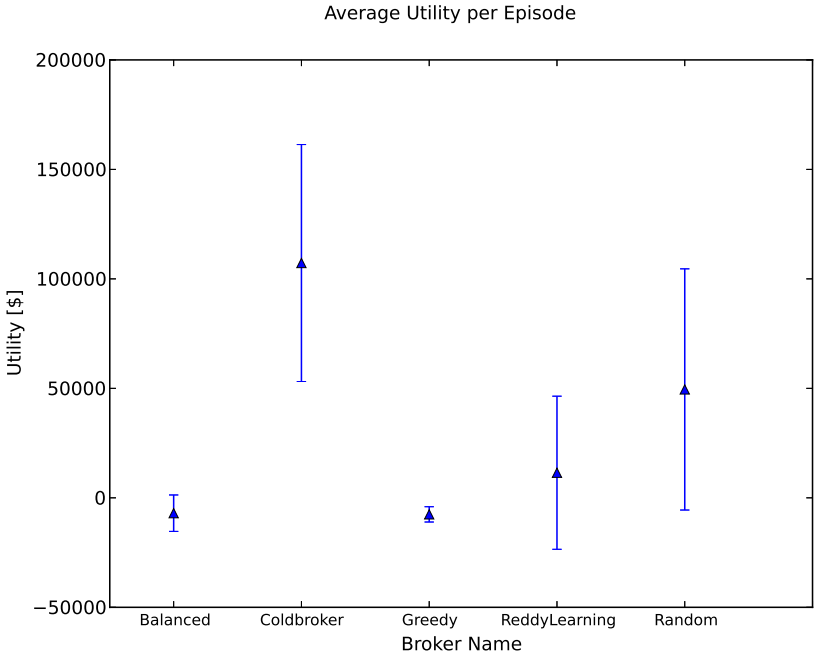

Figure 5.1: Overall average and standard deviation for each broker

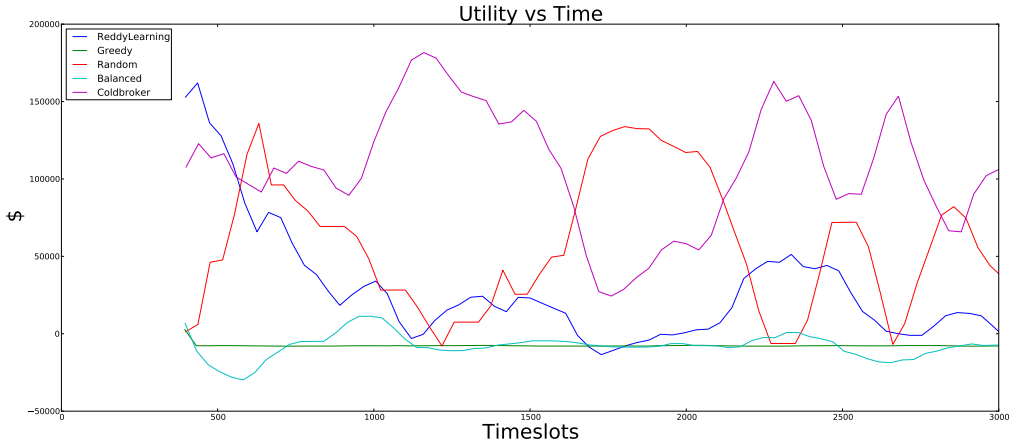

Figure 5.2: Utility for all brokers

EXPERIMENT ANALYSIS

Several observations can be done from this experiment's results. First of all Fig. 5.2 and Fig. 5.1 clearly show that COLD broker has the highest utility compared to the rest of the competing brokers. The second position is for the Random broker and the third one for ReddyLearning. The latter broker has a different set of actions and a different representation of the market, compared to COLD broker and Random. On the other hand, Random shares the same set of actions and the same market representation with COLD broker, for this reason Random gets a better utility that COLD broker sometimes, when it reacts after COLD broker has published its tariffs. This fact highlights the importance of the proposed representation.

It is important to mention that COLD broker's actions are market-bounded, which means that the resulting prices will be competitive, thus customers have a higher probability of deciding to suscribe to them. This means that if the competing brokers are publishing consumption prices around the 0.01 - 0.2 $/MWh zone, COLD broker will not publish something that might result unattractive such as a consumption price of 0.35 $/MWh. ReddyLearning's actions are not market-bounded, this means that all of its actions can lead to a price position that is not

attractive at all for the customers; *i.e.* a consumption price near the maximum consumption price. For this reason, the Random broker obtains a better profit compared to ReddyLearning.

Table 5.3 provides more insight on the brokers' behavior. First of all we can notice that, for COLD broker, even if the overall standard deviation is high compared to the overall average, there are states with higher averages and lower standard deviations. The states with larger average rewards are those when PS_t equals to *rational* and when CPS_t equals far or $very far$. This two values for CPS_t are associated with the inline and bottom actions, which safely place the consumption price away from the competitors, making the published tariff attractive to the customers. These states have as well some of the lowest standard deviations, which tells us that this is a consistent desirable state.

The lowest profits are obtained when PRS_t is *inverted*, because this is by itself a bad state, because the consumption prices are below the production prices. COLD broker learned that the actions that yielded the highest profit on most states were *inline* and *bottom*, and these were the two most used actions, followed by *wide*.

5.5.2 COLD broker vs. ReddyLearning

This section describes the results of a second set of experiments, where COLD broker was tested against its direct competitor ReddyLearning alone. Table 5.4 is similar to Table 5.3, but only for COLD broker and ReddyLearning. Fig. 5.3 shows a plot with the average utility and standard deviation for this experiment.

Experiment Analysis

By looking at table 5.4 it is evident that COLD broker achieves better results than ReddyLearning with a small standard deviation. The average utility on this experi-

	Cold		ReddyLearning	
State	**Average**	**Std. Dev**	**Average**	**Std. Dev**
RashFaOu	191,789	14,565		
RashFaNe	187,515	14,986		
RashNeOu	180,001	11,891		
RashFaVe	153,027	-	169,451	46,892
RashOuFa	150,018	16,176	136,896	52,356
RashOuVe			56,763	32,629
Grand Total	182,963	15,948	59,824	36,457

Table 5.4: Utility for COLD broker and ReddyLearning

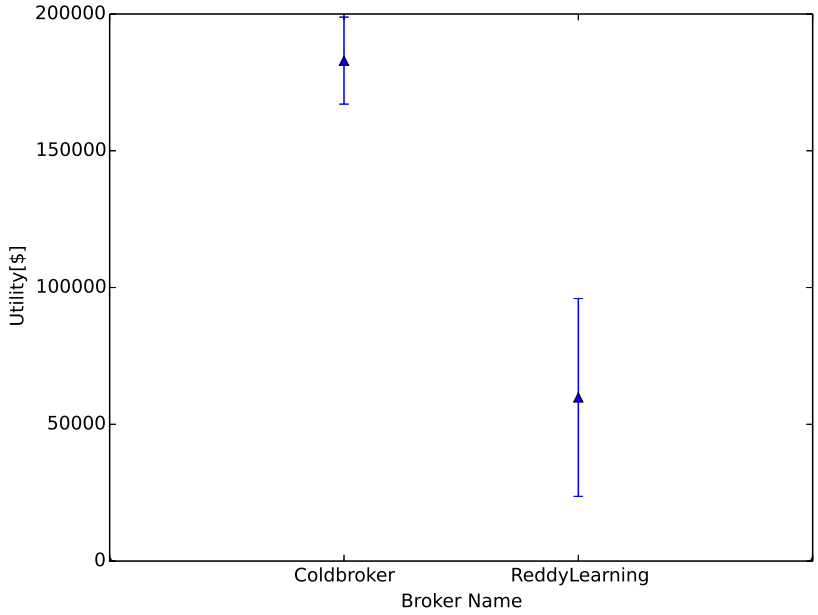

Figure 5.3: Overall average and standard deviation for COLD broker and ReddyLearning

ment compared to Fig. 5.1 is higher, because there are less brokers and for this reason more customers for each one. The fact that the Random broker is not participating has noticeable effects on this experiment:

- The standard deviation is reduced because both learning brokers choose the action that they believe is the one that will yield the highest profit consistently, reducing randomness.
- There are less states because the market state depends on the decisions and

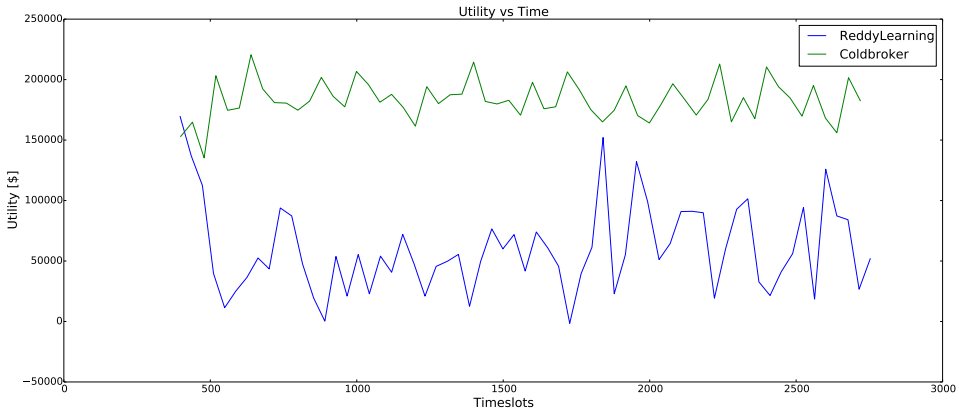

Figure 5.4: Utility obtained when testing COLD broker against ReddyLearning

prices of every broker. When the space of actions is reduced, either because the learning brokers consistently choose the same action or because fixed-strategy brokers choose among one or two actions at most; the list of the different states that the market will achieve is short. However, when the Random broker participates, it chooses among all the available actions with the same probability, creating a wider range of action combinations, and thus, widening the list of different market states.

Table 5.4 offers as well interesting information. The standard deviation columns show that this value is smaller for COLD broker compared to ReddyLearning at every state. COLD broker's values on the Average column are larger for every state compared to those on ReddyLearning; this shows that the COLD broker obtained the largest utility on every decision step consistently. Table 5.4 has as well some similarities with table 5.3; namely the states with the highest average values, as shown on table 5.5.

State *RashFaOu* appears on first place on both experiments, while state RashVeOu appears on second and third place. These states include either the value far or $veryfar$ for CPS_t, which correspond to an expensive consumption price. However,

Position	COLD broker vs. All	COLD broker vs ReddyLearning
1	RashFaOu	RashFaOu
2	RashVeOu	RashFaNe
3	RashFaNe	RashNeOu

Table 5.5: Top 3 states with the highest average profit.

since COLD broker's actions are market-bounded, its most expensive tariff price is still slightly cheaper than the most expensive tariff offered to consumers by any other competing broker. Since the values far and $veryfar$ yield the highest profit for COLD broker, it always executes actions that lead to this value.

The opposite occurs with PPS_t; values *out* and *near* are the lowest production prices that COLD broker can offer, however, these prices are slightly higher that the competitors production prices, making the corresponding tariffs attractive for energy producers.

The value Ra for PRS_t indicates that the market is rational, this is, the consumption prices are above the production prices. If the market was inverted, then none of the brokers would obtain a profit, since on this situation, production prices would be more expensive than consumption prices. For this reason, both learning brokers learned to avoid this state. On this experiment the most used actions were *inline* and *bottom*, followed by *maintain*.

5.5.3 COLD broker vs. COLD broker

The last experiment tested an instance of COLD broker against another instance of itself. The results are very different from those shown in previous experiments. In this ocassion none of the instances is able to obtain a profit, as explained on the next section.

EXPERIMENT ANALYSIS

Fig. 5.5 shows the utility for each instance of COLD broker. Both instances have a peak at the beginning and then decline to reach a minimum, where they stay until the end of the simulation. This occurs because both brokers engage on a price war (also known as arms race in economics) which none of them is capable to win; they respond with the same actions at the same time. As stated before, COLD broker's actions are market-bounded, which means that it will try always to keep its prices between $P_{t,C}^{min}$ and $P_{t,P}^{min}$. However, at each decision step both instances of COLD broker set new values for these two parameters, which tend to be lower that the previous ones. At the end, the market prices move to a point where they are very low and thus, no profit can be obtained by neither of them. This price war can be observed on Table 5.6, which shows on the first column the timeslot where the decision step took place, and on the second and third columns the published consumption tariff price for the first and second instance of COLD broker respectively. The price starts at 0.1333 dollars per MWh and then, with each decision step it is lowered by both brokers, until a bottom limit price of 0.0295 is reached and maintained for the rest of the simulation. On this experiments the most used actions were *inline* and *bottom*, followed by *maintain*.

5.6 GENERAL DISCUSION

In general terms COLD broker achieved a higher utility compared to the other brokers, both against Reddy's learning broker and fixed-strategy brokers. As mentioned before, the Random broker obtained the second largest profit, even higher than that obtained by ReddyLearning. This is a relevant issue and comes from the fact that Random broker uses COLD broker's market representation and set of actions. The reason of why COLD broker outperforms Random is that COLD broker's decissions

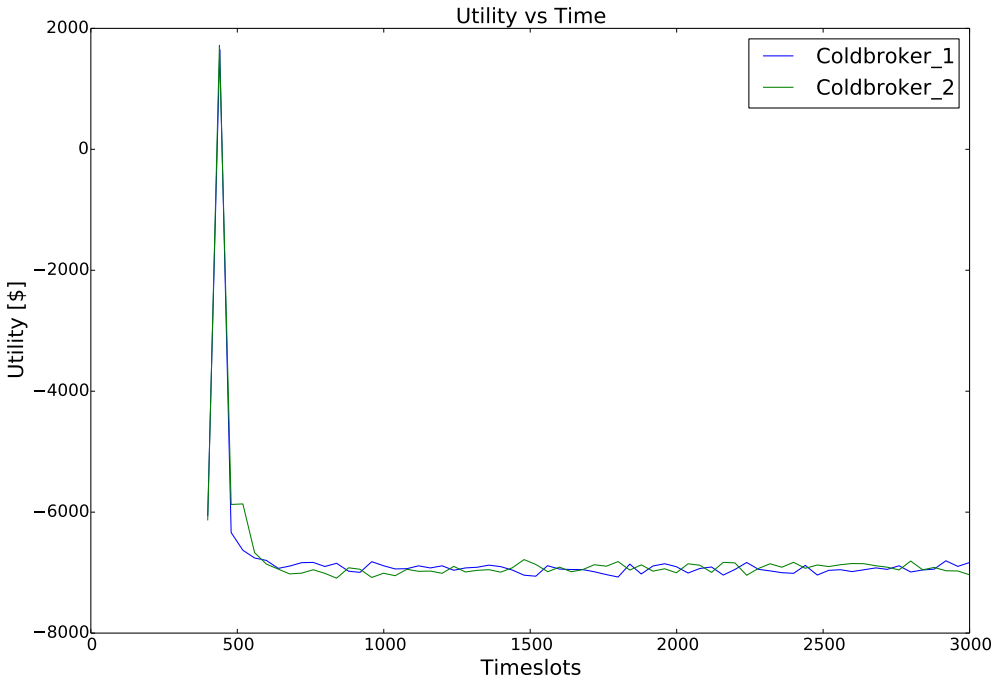

Figure 5.5: Utility obtained when testing COLD broker against itself.

comes from a learning process trained to maximize utility. However, the fact that Random can do a better job achieving utility than ReddyLearning is only related to the proposed set of actions and market representation. If a learning process can't yield a higher utility than a random strategy could only mean that the actions available for the random strategy yield better utilities by themselves than those available for the learning strategy. If we then replace the random strategy by a learning strategy, capable of picking the best action given the market situation, then we achieve the results shown in this section.

Timeslot	COLD broker 1 Consumption Price	COLD broker 2 Consumption Price
399	0.1333	0.1333
440	0.1066	0.1333
479	0.0853	0.1066
519	0.0552	0.0853
559	0.0386	0.0552
599	0.0295	0.0386
640	0.0295	0.0295
679	0.0295	0.0295

Table 5.6: Published consumption prices for the first decision steps for COLD broker 1 and COLD broker 2

Chapter 6

Conclusions

The non-renewable characteristic of oil is driving without a doubt to seek efficient and sustainable alternatives to satisfy the world's energy demand. Centralized energy monopolies like the ones existing on most world's countries will reach a limit as the energy demand keeps increasing and as the oil production declines. An alternative to this scenario is the distributed and self-sustained smartgrid, where buildings, houses and even communities are able to produce, by renewable means such as solar panels and wind turbines, the electricity they require; and sell to any other instance the energy they will not use. This smartgrid will provide, on the long term a cheap and almost limitless amount of energy, and the means to efficiently distribute the energy where it is required.

However, the smartgrid faces hard challenges, and beyond those related to the technology to physically integrate the encoders, receivers and transmitters, which will allow the smartgrid to work as a communications channel; there exists challenges related to the algorithms and software that will handle and process the received signals. These signals will be used by a regulatory entity to ensure that the system will remain balanced, and will be used by the trading brokers to determine which features a tariff requires to satisfy customers demands. There is still a lot of work to be done to fully understand the behavior of a smartgrid, and how a broker's decisions can impact the energy flow, or the whole smartgrid stability. For this reason simulating platforms such as Power TAC are useful; it allows scientists to

propose and safely test energy trading algorithms and rules, and, as described before on this thesis work, one of the main energy trading scenarios is the tariff market.

A tariff may have several attributes, but one of the most important is the price. It is not an easy task to determine the proper pricing for a tariff at a given time, because the utility the broker might expect for choosing any action, and thus determining a tariff's price, depends on the market state; and the market state depends on the actions of all the brokers. For this reason, the most important step that can be taken, is to provide a market model which captures its main attributes, to be used as the input to the decision algorhitm. This thesis proposed a market model different to the one proposed on previous publications. Also, if the broker has a good market representation, but lacks of a set of good actions to respond to the market, its performance will be poor in terms of utility. For this reason, this thesis also proposed a set of actions which adapts its price output to the changing market states.

The experiments showed that COLD broker, with its proposed set of actions and its market representation was able to obtain the highest profits on 70% of the evaluated timeslots, when tested against all the competing brokers, including ReddyLearning. When tested only against the latter, COLD broker was able to obtain the highest profit 100% of the evaluated timeslots. This proved that both the market representation and the proposed actions achieved a better average utility compared to that delivered by the algorithms found on related previous works, and used by the other competing brokers.

It is important to mention as well that the market representation size is not bounded to the number of competing brokers; the number of possible value combinations of state space S will remain the same if there are 1, 2 or more competing brokers. This is very useful because it makes easier the learning process. On the other hand, the proposed market-bounded actions were the most used by COLD

broker, and these actions conducted it to lead the utility rank most of the time on the experiments executed. Even as there were some non-market-bounded actions available, such as *Minmax* for instance, COLD broker learned that those actions did not yield good results, and for this reason decided not to use them.

There are interesting works that can be done to further improve the research done on this thesis. The future work can be related to any of these aspects.

- Tariff characteristics: this thesis experimented only with various price configurations for a consumption and production tariff. However a tariff has plenty of other attributes that customers consider while evaluating a tariff's utility. Attributes such as signup payments, early withdraw payments or complex rate tariffs such as block or time-of-use tariffs might be explored and included on a learning algorithm to further improve the broker's utility.

- Number of tariffs selection: the experiments conducted in this thesis included only one consumption and one production tariff at any given timeslot; so this is an area that might be further explored to determine how many tariffs must be available for customers to suscribe.

- Tariff publication and revocation timing: the broker we developed publishes its tariffs at a fixed interval. However, developing a strategy to determine when a tariff is to be published or revoked could lead to a creating a strong customer base for the broker.

- Tariff to revoke selection: another related issue is to select which tariff has to be revoked. A naive strategy might be to revoke, among the broker's tariffs portfolio, the tariff with the lowest profit. However other strategies might lead to keep tariffs that may not yield large profits short after being published, but would pay out on the long term.

- Learning procedure: this thesis used Q-Learning as the RL method to determine the action to be executed at each evaluation period. A drawback of this method is that it requires several iterations, to acquire enough knowledge about the environment, before it can actually be profitable for the learning broker. This can lead to enormous losses if the learning process occurs online. An alternative to this is to learn offline and then applied the acquired knowledge against real competitors. By doing this, a trade is done between learning and adaptability. If the learning process is done offline with a given set of competitors, and then tested online; this testing process has to include the same set of competitors to be effective. If the competitor set is different, the acquired learning may not be very effective. This scenario is not far away from reality, where the algorithms to generate tariffs or the brokers generating the competitors tariffs are not known during the training process, but until a competition is set. For this reason, it might be useful to experiment with other learning algorithms which can learn faster at a lower cost.

Appendix A

Details of the Analysis on Rationality Parameter Lambda

As stated on Sec.2.4.3 the parameter λ determines the customer's rationality. Customers evaluate a tariff's T_a utility by comparing it with the utility of a reference tariff T_r. Assuming flat tariffs, as the price of T_a moves away from that of T_r, the utility of the evaluated tariff increases. However, the rationality parameter has the power to influence on this perception. If the rationality is high, even a small difference between the price of T_a and T_r will consistently yield a high utility (let us call it U_a) for T_a. If the rationality decreases, then a larger difference D between the prices will be required to achieve the same value of T_a. If the price difference is less than D, customers will not be able to tell any difference between such tariffs and therefore will choose randomly between them. This is important because it means that on highly rational markets, it is possible to obtain greater utilities by charging higher prices, as long as these prices are slightly below the reference price for consumption tariffs, and slightly above the reference price for production tariffs. On markets with a low or average rationality (as the one modeled by Power TAC) the most attractive tariff prices will not be as close to that of T_r, making it harder to determine the correct ones.

In order to explore the rationality parameter λ some experiments were con-

ducted. The experiments included a series of simulations where a consumption and a production tariff were published at the beginning, and then left unchanged until the simulation ended. The accumulated utility was observed at fixed intervals within each simulation, then it was recorded and reseted. At the end of each simulation the average, and its corresponding standard deviation, was calculated from the observation recordings.

Several pairs of prices were published in the same way: publishing them at the beginning and keeping them unchanged, measuring the average and standard deviation of the recorded accumulated utilities, until another pair was tested. The first set of experiments were conducted with $\lambda = 5$. Then, a second set of experiments was conducted with identical production/consumption pairs, but now with $\lambda = 1000$. According to Sec. 2.4.3 we expected to obtain consistently a greater profit on the experiments with the larger λ, compared with those tested with the lower λ. The recorded standard deviation allowed us to measure how consistent this lower or higher profits were. Figures A.1 and A.2 show the results of this series of experiments. The specific settings of the experiment are described as follows.

Two brokers were instantiated to compete over 34 different simulations, each with a length of 3,000 timeslots. From this set, 26 of experiments were tested with a low value for λ and 8 of them were tested with a high value for λ. The brokers published a pair of initial tariffs, one consumption tariff and one production tariff. After publishing these two fixed-rate tariffs, they were not changed until the simulation ended, following the process described in the previous paragraphs. One of the brokers published the reference consumption and production tariffs with a value of 0.5 and 0.015 respectively, while the other broker published the different combinations of production and consumption tariffs, as described by Table A.1. The accumulated utility was measured at the end of 40 timeslot intervals.

In Table A.1 the consumption prices are the columns and the production prices

are rows. At each row-column interception there is either an x, which indicates that an experiment was conducted with that combination of production and consumption prices, or a blank space. Blank spaces indicate that there was no experiment conducted with that combination.

P/C	0.500	0.425	0.350	0.275	0.240	0.165	0.090	0.015
0.240	x	x	x	x	x			
0.165	x	x	x	x	x	x		
0.090	x	x	x	x	x	x	x	
0.015	x	x	x	x	x	x	x	x

Table A.1: Consumption and production prices combinations.

The plots on figures A.1 and A.2 show the results obtained with these experiments. Figure A.1 shows the average of the accumulated utilities as observed at the end of each 40-timeslot evaluation periods. The horizontal axis shows the consumption prices, and the series on the box show the production prices. The corresponding standard deviation for each of the experiments is plotted on figure A.2. The experiments with the $\lambda = 5$ (low rationality) are plotted with the continuous lines, and the experiments with $\lambda = 1000$ (high rationality) are plotted with the dotted line.

LAMBDA EXPERIMENTS ANALYSIS

The experiments showed that, with $\lambda = 5$ (continuous plots), the highest utilities are reached when the consumption price is around 0.275, regardless of the production price. As the consumption price increases towards 0.5, the average utility is reduced and the standard deviation peaks. The high standard deviation values near the reference consumption price at 0.5 indicate that customers do not always assign a high utility to tariffs around this area, regardless of the associated production price.

When the rationality is high (dotted plots), the utility reaches a peak when the

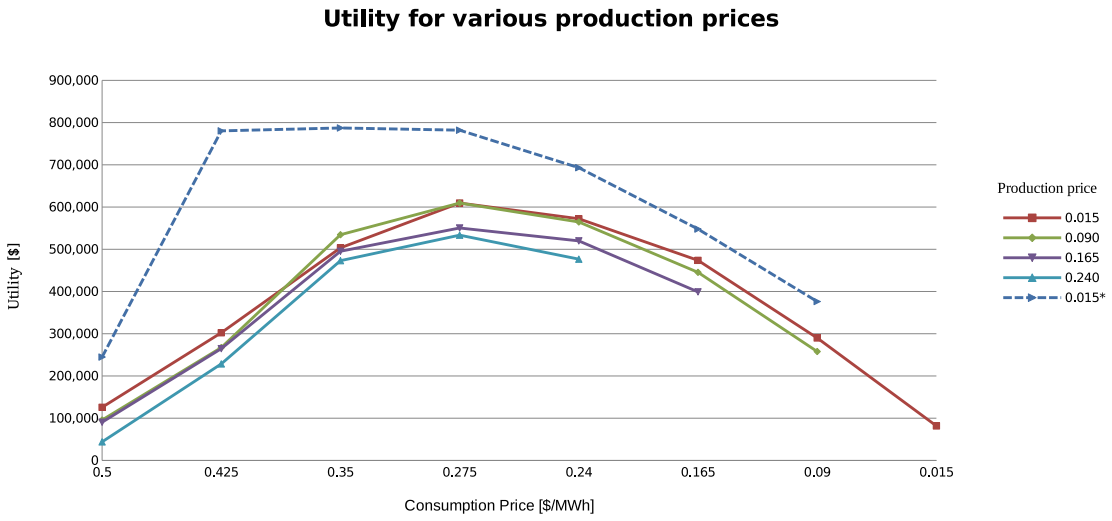

Figure A.1: Customer's rationality effect on broker's utility

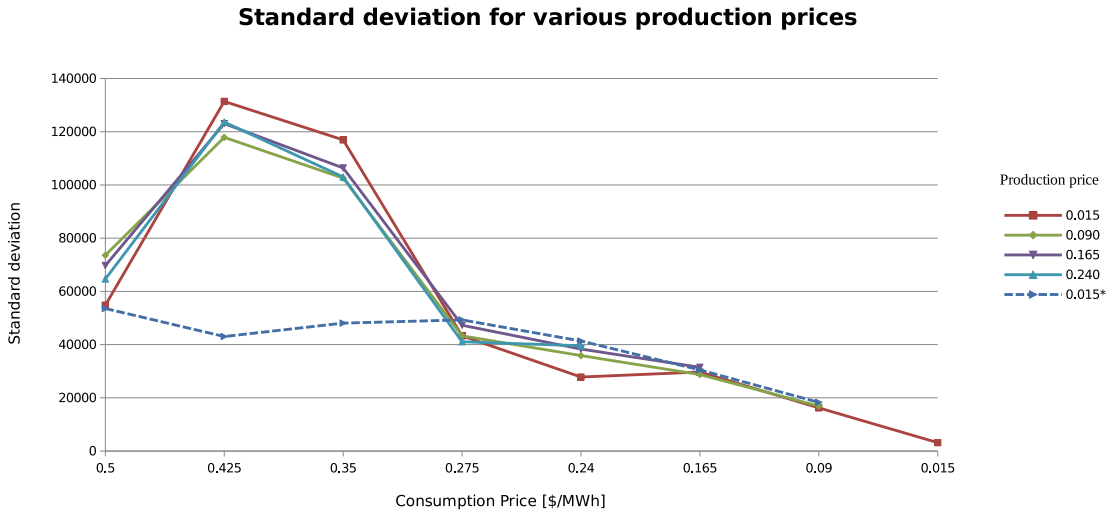

Figure A.2: Standard deviation values for each of the conducted experiments.

consumption price is around 0.450, while the standard deviation reaches a maximum around 0.275 and then keeps constant towards 0.500. This effect matches what was expected at the beginning of this section: on a highly rational market, a broker will have higher utilities by publishing expensive consumption tariffs and cheap production tariffs.

Appendix B

Example of Data Book Keeping

COLD broker's maintains a file with log containing relevant information. This information is stored in a file called States. This file stores the data on various columns. Each register includes information about the timeslot, the published consumption and production prices, the action performed, the MDP state and the reward obtained.

Table B.1 shows an example of the previously mentioned log. The first column indicates the timeslot when the decision was taken, columns 2 to 5 show the price top and bottom values, columns 6 to 9 show the values that determine the market state, column 7 shows the action performed and column 8 shows the utility achieved. So, for instance at timeslot 479 in Table B.1, an action bottom was executed, which generated an utility of 135,251 and set the market state to RaShNeOu on timeslot 519.

t	$P_{t,C}^{max}$	$P_{t,C}^{min}$	$P_{t,P}^{max}$	$P_{t,P}^{min}$	PRS_t	PS_t	CPS_t	PPS_t	Action	Utility
399	0.5000	0.1311	0.0508	0.0150	Ra	Sh	Fa	Ve	ma	153,027
439	0.5000	0.1020	0.0798	0.0150	Ra	Sh	Ou	Fa	ma	164,785
479	0.5000	0.1020	0.0798	0.0150	Ra	Sh	Ou	Fa	bo	135,251
519	0.5000	0.1020	0.0798	0.0150	Ra	Sh	Ne	Ou	bo	203,226
559	0.5000	0.1020	0.0798	0.0150	Ra	Sh	Ne	Ou	in	174,601
599	0.5000	0.1020	0.0798	0.0150	Ra	Sh	Fa	Ou	in	176,479
639	0.5000	0.1020	0.0798	0.0150	Ra	Sh	Fa	Ou	bo	220,555
679	0.5000	0.1020	0.0798	0.0150	Ra	Sh	Ne	Ou	bo	192,274
719	0.5000	0.1020	0.0798	0.0150	Ra	Sh	Ne	Ou	bo	180,977
760	0.5000	0.1020	0.0798	0.0150	Ra	Sh	Ne	Ou	bo	180,551

Table B.1: An example on how COLD broker keeps a record of its published prices and utilities.

Bibliography

[Babić, 2012] Babić, J. (2012). *A simulation platform for power trading*. PhD thesis, Master Thesis, University of Zagreb, Faculty of electrical engineering and computing.

[Barbose et al., 2004] Barbose, G., Goldman, C., and Neenan, B. (2004). A survey of utility experience with real time pricing. *Lawrence Berkeley National Laboratory*.

[Barto, 1998] Barto, A. G. (1998). *Reinforcement Learning: An introduction*. MIT press.

[Ben Elghali et al., 2007] Ben Elghali, S., Benbouzid, M., and Charpentier, J. F. (2007). Marine tidal current electric power generation technology: State of the art and current status. In *Electric Machines & Drives Conference, 2007. IEMDC'07. IEEE International*, volume 2, pages 1407–1412. IEEE.

[Borenstein, 2009] Borenstein, S. (2009). To what electricity price do consumers respond. *Residential Demand Elasticity Under Increasing-Block Pricing. Berkeley, CA*.

[Borenstein et al., 2002] Borenstein, S., Jaske, M., and Rosenfeld, A. (2002). Dynamic pricing, advanced metering, and demand response in electricity markets. *Center for the Study of Energy Markets*.

[Farahmand et al., 2012] Farahmand, H., Member, S., Aigner, T., Member, S., Doorman, G. L., Member, S., Korpå s, M., and Huertas-hernando, D. (2012).

Balancing Market Integration in the Northern European Continent : A 2030 Case Study. 3(4):918–930.

[Farhangi, 2010] Farhangi, H. (2010). The path of the smart grid. *Power and Energy Magazine, IEEE*, 8(1):18–28.

[Faruqui, 2010] Faruqui, A. (2010). The ethics of dynamic pricing. *The Electricity Journal*, 23(6):13–27.

[Gordijn and Akkermans, 2007] Gordijn, J. and Akkermans, H. (2007). Business models for distributed generation in a liberalized market environment. *Electric Power Systems Research*, 77(9):1178–1188.

[Haas et al., 2011] Haas, R., Panzer, C., Resch, G., Ragwitz, M., Reece, G., and Held, A. (2011). A historical review of promotion strategies for electricity from renewable energy sources in eu countries. *Renewable and Sustainable Energy Reviews*, 15(2):1003–1034.

[Ilie et al., 2007] Ilie, L., Horobet, A., and Popescu, C. (2007). Liberalization and regulation in the eu energy market. Technical report, University Library of Munich, Germany, Munich Personal RePEc Archive.

[Ipakchi and Albuyeh, 2009] Ipakchi, A. and Albuyeh, F. (2009). Grid of the Future. *Power and Energy Magazine, IEEE*, 7(2):52–62.

[Joskow and Tirole, 2006] Joskow, P. and Tirole, J. (2006). Retail electricity competition. Technical Report 4.

[Keppo and Räsänen, 1999] Keppo, J. and Räsänen, M. (1999). Pricing of electricity tariffs in competitive markets. *Energy economics*, 21(3):213–223.

[Ketter and Collins, 2013] Ketter, W. and Collins, J. (2013). The 2013 Power Trading Agent Competition. (May).

[Ketter et al., 2013] Ketter, W., Collins, J., Reddy, P. P., and Weerdt, M. D. (2013). The 2013 power trading agent competition. *ERIM Report Series Reference No. ERS-2013-006-LIS*.

[Kirschen, 2003] Kirschen, D. S. (2003). Demand-side view of electricity markets. *Power Systems, IEEE Transactions on*, 18(2):520–527.

[Lund et al., 2012] Lund, H., Andersen, A. N., Alberg, P., Vad, B., and Connolly, D. (2012). From electricity smart grids to smart energy systems e A market operation based approach and understanding. *Energy*, 42(1):96–102.

[Maenhoudt and Deconinck, 2010] Maenhoudt, M. and Deconinck, G. (2010). Agent-based modelling as a tool for testing electric power market designs. *2010 7th International Conference on the European Energy Market*, pages 1–5.

[NIST, 2012] NIST (2012). Smart grid: a begginer's guide. `http://www.nist.gov/smartgrid/beginnersguide.cfm`. Accessed: 2014-05-10.

[North et al., 2002] North, M., Conzelmann, G., Koritarov, V., Macal, C., Thimmapuram, P., and Veselka, T. (2002). E-laboratories: agent-based modeling of electricity markets. In *2002 American Power Conference*, pages 15–17.

[Puterman, 2005] Puterman, M. L. (2005). *Markov Decission Processes*. Wiley, New Jersey, 3 edition.

[Reddy and Veloso, 2011] Reddy, P. P. and Veloso, M. M. (2011). Learned behaviors of multiple autonomous agents in smart grid markets. In *AAAI*.

[Ringel, 2006] Ringel, M. (2006). Fostering the use of renewable energies in the European Union: the race between feed-in tariffs and green certificates. *Renewable Energy*, 31(1):1–17.

[Watkins and Dayan, 1992] Watkins, C. J. and Dayan, P. (1992). Q-learning. *Machine learning*, 8(3-4):279–292.

Glossary

broker a trading agent who offers electic tariffs to either consumers or producers, with the purpose of obtaining a profit. 2

customer a market entity who is suscribed to one or more broker'r tariffs and either consumes or produces energy. 4

decission step is the instant where the broker evaluates the available information about the market state and chooses an action. 43

electricity commodity is a marketable item produced to satisfy wants or needs. Electricity has the particular characteristic that it is usually uneconomical to store; hence, electricity must be consumed as soon as it is produced. 6

evaluation period see decission step. 43

imbalance the difference between the energy sold and acquired by a broker. 18

Power TAC is an open source smartgrid simulation platform developed on Java. 19

regulated distribution utility a centralized entity which owns the distribution lines and is in charge of regulating and applying charges to brokers incurring on inbalance. 18

retail markets the market where the tariff contracts are traded. The main characteristic of this market is that there is a large volume of transactions, each trading small amounts of energy. 18

risk premium the return in excess of the risk-free rate of return that an investment is expected to yield . 6

smartgrid is a modernized grid that enables bidirectional flows of energy and uses two-way communication and control capabilities that will lead to an array of new functionalities and applications. 1

tariff is an agreemen which grant customers, either producers or consumers, the right to trade with a broker certain amount of energy under . 20

tariff contracts see tariff. 18

wholesale market the market where brokers can sell and acquire energy in large volumes. The main characteristic of this market is that there is a low volume of transactions, each with trading large amounts of energy. 18