Luis Briones Roselló

Market-based control of plug-in electric vehicles

Master Thesis
SS2013

EEH – Power Systems Laboratory
Swiss Federal Institute of Technology (ETH) Zurich

Supervisor: Prof. Dr. Göran Andersson
Assistant: Marina González Vayá

Zurich, September 2, 2013
Abstract

The integration of a high number of plug-in electric (PEV) vehicles could lead to overloads in the systems assets and demand peaks if the charge of the fleet is left uncontrolled. However, with the use of smart-charging strategies these problems could be avoided. In this work the development of a smart-charging strategy is presented. The goal of each electric vehicle, modeled as an agent, is to minimize the cost of energy purchase while satisfying the energy requirements. To solve this problem, multi-agent system theory is used in combination with market-based control. The vehicles are considered as agents bidding on the market, optimizing their bidding to minimize their costs. An aggregator agent acts as communication middleman between the vehicles and the market. This way, a system with a high number of agents competing for the resources is established. The resources are allocated according to the demand-supply theory, and the equilibrium price of the day-ahead market is used as a control signal. Moreover, a Q-learning algorithm is used for the learning process of the vehicles, establishing their optimal bidding strategy. In our case studies, we analyze this approach both in a simple market clearing and an Optimal Flow setting. Moreover, we analyze the effect of uncertainties in driving patterns and non-PEV bids. The results show that the use of this strategy leads to a lower energy costs for the vehicles. The fleet charges mainly during the night hours, avoiding the charge during demand peaks.
Acknowledgements

I would like to thank my supervisor Marina González Vayá for the support and patient guidance during the development of this thesis. Thanks to the Prof. Dr. Göran Andersson for giving me the opportunity to write my thesis on the Power system laboratory at ETH Zürich. Furthermore a thank-you goes to all the people I have met here in Zürich this last six months. I had a great time here, and it would not have been the same without you. Last, but not least, I cannot forget about all my friends and colleagues from Valencia, and specially my family for the encouragement they gave me during my studies.

Zürich, September 2, 2013

Luis Briones Roselló
Contents

Abstract ii

List of Acronyms ix

List of Figures x

1 Introduction 1

2 Multi-agent Systems 5
   2.1 Multi-Agent System 5
   2.2 Market based Control 6
   2.3 Privacy 7

3 Market Clearing 9
   3.1 Electricity Market 9
      3.1.1 Copper plate market clearing 10
      3.1.2 Optimal Power Flow 10

4 Reinforcement Learning: Q-Learning 13
   4.1 Q-learning algorithm 14
   4.2 The $\epsilon$-Greedy policy 14

5 The Market-based control model 17
   5.1 PEV agents 18
      5.1.1 Definition of the Demand Bids 18
      5.1.2 Actions 21
   5.2 Aggregator agent 22
      5.2.1 Aggregating process 22

6 Implementation 25
   6.1 General implementation 25
   6.2 Implementation of the Q-learning 26
# Contents

7 Results

7.1 Base case .................................................. 29
7.1.1 Results and comments ................................. 29

7.2 Integration into an OPF market clearing .......... 32
7.2.1 Results and comments ................................. 33

7.3 Uncertainty in the behavior of the vehicles ....... 35
7.3.1 Results and comments ................................. 36

7.4 Uncertainty in demand supply ......................... 38
7.4.1 Results and comments ................................. 38

8 Conclusions .................................................. 41
8.1 Outlook ...................................................... 42

Bibliography .................................................... 43

Appendices ..................................................... 44

A Definition of the actions ................................. 45
A.1 Definition of $D_p$ .................................. 45
A.2 Definition of $C_p$ .................................. 46

B Q-learning parameters .................................... 49
B.1 Learning rate and $\epsilon$-Greedy policy .......... 49
B.2 Penalty term ........................................ 51

C Matlab code .................................................. 55
List of Acronyms

PEV  Plug-in Electric Vehicles (denotes both plug-in hybrid and pure electric vehicles)

MAS  Multi-Agent System

MBC  Market-Based Control

OPF  Optimal Power Flow

PTDF  Power Transfer Distribution Factors
List of Figures

5.1 General Scheme ................................................. 18
5.2 Demand bid ......................................................... 20
5.3 Individual and aggregated demand curves ........................... 23

6.1 Flow diagram of the daily simulation ............................... 26

7.1 Base case: System load, market prices and charging profile of the fleet ......................................................... 30
7.2 Base case: Evolution of the cost of the energy ............... 31
7.3 Base case: Total reward for the whole fleet .................. 32
7.4 Base case: actions performed by one of the vehicles during the simulation ......................................................... 33
7.5 Integration into an OPF: System load and charging profile of the fleet ......................................................... 34
7.6 Integration into an OPF: Nodal prices in the scenario without PEVs ......................................................... 34
7.7 Integration into an OPF: Nodal prices in the scenario with inflexible demand of the PEVs ................................................. 35
7.8 Integration into an OPF: Nodal prices in the scenario with the use of the bidding strategy ................................................. 35
7.9 Vehicles behavior uncertainty: System load, market prices and charging profile of the fleet ......................................................... 36
7.10 Vehicles behavior uncertainty case: Evolution of the cost of the energy ......................................................... 37
7.11 Vehicles behaviour uncertainty case: Total reward for the whole fleet ......................................................... 37
7.12 System load, market prices and charging profile of the fleet for the week June 4th-June 10th, 2012 ......................... 39

A.1 Clearing price of the market without vehicles ............... 46
A.2 Base case: System load, market prices and charging profile of the fleet with two different sets of actions ................. 47
A.3 (a) Distribution of the minimum clearing prices of the market without vehicles in 2012 (b) Detail of (a) .................. 48
| B.1 | Base case: System load, market prices and charging profile of the fleet with different $\epsilon$-Greedy policies and learning rate values | 50 |
| B.2 | Base case: Reward values with different $\epsilon$-Greedy policies and learning rate values | 50 |
| B.3 | Base case: Actions performed by one of the vehicles during the simulations with different $\epsilon$-Greedy policies and learning rate values | 51 |
| B.4 | Integration into an OPF: System load and charging profile of the fleet with different penalty policies | 52 |
| B.5 | Integration into an OPF: Actions and Reward with different penalty policies | 53 |
Chapter 1

Introduction

Environmental issues as well as political instability in the oil producing countries are the main reasons to move from a fossil fuel dependency model for the transportation sector to an integration of both plug-in hybrid and pure electric vehicles, which we will denote in the following as PEVs. PEVs are seen as a key to reduce that dependency, as well as the greenhouse gases emissions. Moreover, they could play an active role in the integration of renewable energy resources into the grid. However, if the charging process of a great number of vehicles is left uncontrolled, that could lead to a situation with demand peaks or overloads in system assets. As flexible loads, the demand of PEVs can be shaped with smart charging strategies. Those strategies should be performed to avoid the problems brought about by uncontrolled charging.

Here, the implementation of a smart-charging strategy based on the cost minimization of the energy purchased is proposed. More precisely, the goal of the thesis is the development of a market-based control for the charging of plug-in electric vehicles.

Two examples of cost-minimizing charging strategies are the centralized and decentralized approaches.

In a centralized approach the scheduling of the charge is determined by a central agent. It should determine a charging profile that satisfies the demand and minimizes the cost. In a decentralized approach, PEVs are able to decide when to charge. Each vehicle tries to optimize its own cost.

Both schemes can show good results in terms of generation cost and network assets usage. For instance, in [1], the centralized approach leads to a least-cost solution, but the decentralized scheme results are comparable.
One of the advantages of decentralized schemes is that they have lower communication requirements. It is sufficient to broadcast a signal, e.g. prices, to control the charge. However, the possibility of a simultaneous reaction (avalanche effect) or forecasting errors of the consumer reaction to the price signals are disadvantages of this approach.

By contrast, an important feature of the centralized approach is that it takes into account the effect of the PEV fleet on the prices. In order to take part in the daily market, the required information of the PEVs has to be sent to the central level. It schedules the charge, and the information returns to the vehicles. Thus, an inconvenient of a centralized approach is the required bidirectional communication.

In this thesis an intermediate scheme is studied. A multi-agent system is established, where each vehicle tries to optimize its own cost, as in a decentralized approach. In this case, each PEV states its willingness to purchase energy with the placement of demand bids, defined according to status variables of the vehicle. Those bids are aggregated in an intermediate level, and sent to a central level, where the market clearing takes place.

As the vehicles participate as loads on the market, their influence on the prices is taken into account. Moreover, vehicles respond to the market price according to their bids. Thereby possible avalanche effects are avoided, because the prices are determined in an endogenous way.

The disadvantage of this system, as in the centralized scheme, is that it requires a bidirectional communication. However, the only information required from the vehicles are the bids and the only information sent to the vehicles is the equilibrium price of the market, which is used as the control signal. That way, the system requires a limited amount information to be shared. The role of the aggregator agent is important for the scalability of the communications.

A well-known example and widely developed tool to manage systems with a high number of agents is PowerMatcher. In [2] and [3] the concept is exposed. PowerMatcher is a “market-based control concept for supply and demand matching in electricity networks ” [2]. Based on Multi-Agent system theory, a structure with different communication levels between agents is established.

The main actors of the scheme are Device agents, Concentrator agents and the Auctioneer agent. The communication between agents is based on the exchange of bids and price data. The Device agents, representing the supply-demand elements, place bids. Those bids express the willingness of
buying or selling electricity. The Auctioneer agent concentrates the bids, and searches for the equilibrium price. That price is communicated to the devices, and it determines the amount of power that the agent should produce or consume. Concentrator agents are intermediate agents. They aggregate the bids of the devices, and communicate to the superior level, and they send the updates of the price data to the different device agents.

An important aspect addressed in this thesis is how the agents set up their bids optimally. An interesting description of the definition of the bids and the aggregation process is presented in [4]. The demand curve of the cars is built based on status variables, such as the state of charge of the batteries or the time to departure of the vehicles, and is performed as a piecewise linear function. The aggregation process is based on a tree structure, with similar agents as in the PowerMatcher approach.

In the scheme developed in the thesis, vehicles behave as intelligent agents who try to minimize their own cost. Each vehicle should learn how to place optimal bids changing the parameters used to define the bids based on vehicles status. With this purpose a machine learning algorithm is used. In particular, a reinforcement learning algorithm, called Q-learning is implemented. Some examples of the application of this algorithm in electricity markets are found in [5] and in [6].

The scheme here developed is tested in four different cases. The main goal of the base case is to set up the parameters of the algorithm and, at the same time, study the performance of the control system. The results of this first case show that the bidding strategy implemented leads to a cost reduction of the energy purchased. Moreover, the demand of the vehicles is shaped, and the vehicles charge mainly during the night hours, avoiding the demand peaks and charging when the prices are lower.

In the base case the network constraints are not taken into account. For that reason, the second case study is the integration into an optimal power flow. The approach shows good results in terms of shaping the demand of the fleet and avoiding the charging during the peaks. Moreover, the nodal prices with the bidding strategy here implemented are lower than in the case of an uncontrolled charge. This result is good both in terms of cost minimization and congestion of the network.

The other cases study vary the base case introducing some uncertainty in the input data. In the third case the uncertainty in the behavior of the fleet is introduced. The demand of the vehicles is mainly displaced to the night hours as well.
The fourth case introduce uncertainty in the demand-supply side. Real data from EEX for the years 2011 and 2012 in Germany and Austria is used in this case. The results show that the bidding strategy used here is able to shape the demand and, at the same time reduce the cost of the energy, even when market conditions are constantly evolving.
Chapter 2

Multi-agent Systems

The traditional management of the electrical infrastructure is changing due to the integration of new active elements. The idea of a top-down operation, where the demand is considered inelastic and the supply is dispatched to match with it, is being challenged. New sources such as photovoltaic, wind power or small CHP plants pose the need to develop a novel operation.

Features as intermittent or uncertain generation, depending on the weather, as well as the interaction with a huge number of decentralized supply-demand users, e.g. PEVs, requires a new operation model. To achieve an efficient use of the generation and grid assets, the concept of the inelastic demand needs to be reconsidered. Matching demand with real time supply will require intelligent systems to manage and control this complex infrastructure. Multi-Agent Systems (MAS) are well-suited for this purpose. Moreover, the coordination of a MAS with a Market Based Control contribute to a better decision making process of the different agents.

In this chapter, an introduction to the main concepts of a Multi Agent system will be developed. The definition of the different elements and the operation of such a system will be presented, as well as the use of a Market-Based Control strategy.

2.1 Multi-Agent System

Multi-Agent Systems are used in a wide range of fields, such as power system restoration, automation, network control, or market simulation. The reason is that they allow the management of complex systems by modeling them as groups of intelligent agents that interact with their environment, and are able to change their behavior accordingly.
CHAPTER 2. MULTI-AGENT SYSTEMS

According to Wooldridge[7], agent, environment and autonomy are the main concepts to define a Multi-Agent System.

Agents are the different entities of the system. They could be software or hardware agents. The task of each agent is to act as a representative of something or someone (e.g. the PEVs). The concept of environment refers to all the external elements to an agent. The agent interacts with the environment, and is able to alterate it. The environment should be measurable in order to asses the effects of the agent’s actions. The autonomy of the agent should allow him to adapt and answer to the different environment states.

Finally, in the previously mentioned paper also a differentiation between the concepts of agent and intelligent agent is done. The difference between them is that an intelligent agent has a flexible autonomy, i.e. is able to react to changes, can interact with other intelligent agents and have goal-directed behavior.

According to the previous definitions, the concept of a MAS refers to a system comprising two or more agents or intelligent agents. One interesting feature is that there is no system goal, but only the different individual goals of each agent.

An application example of MAS in electricity networks is found in [2]. It is presented as a solution for the management of a network with high integration of DG. An important idea is that while the whole scheme represents a very complex system, with a high level of intelligence, the complexity of the individual agents remains low. In addition, the decision making process is discussed. As stated on the paper, the interaction of a MAS with an electronic market can make that process more efficient. That way, the basis for a Market-Based Control is set.

2.2 Market-based Control

Market-Based Control (MBC) is a control strategy based on the economic model of supply and demand for price determination. All the agents compete for the resources on the market by allocating bids. The resource assignment is determined by the equilibrium price of the market.

In [8] a market-based mechanism for scheduling short term energy consumption is presented. In this approach, a scheme with an intermediate aggregator agent between the central market and the individual agents is used. The use of a market-based control is justified by the high number of
distributed agents. It is proven that this control allows the management of the decentralized resources.

Another example is [2]. In this case there is a scenario with multiple agents with different behaviors. Stochastic devices, shiftable operation devices or electricity storage devices, among others, are the agents of this cluster. The combination of the MAS with the MBC is used to control the agents. It is shown that the results are similar to a centralized optimizer where the information is submitted to the central level. In terms of scalability, MBC has lower communication requirements. Relevant information, i.e. the bids of the agents, is gathered in order to achieve the optimization. The equilibrium price of the market is the control signal.

2.3 Privacy

One of the key issues in the development of a smart grid system is the privacy. Even though it is not one of the topics developed in the thesis, it is interesting to comment it.

The bidding and communication structure presented has some advantages in terms of security.

A bidirectional information flow is establish for the communication between the different agents. Demand bids are the only information shared by the consumer. They represent the amount of energy and the price the user is willing to pay for that energy. Although bids are built using state parameters of the car, for instance the required state of charge of the battery or the expected time to departure, personal information is not included in the bid. That way no compromising information is shared.
CHAPTER 2. MULTI-AGENT SYSTEMS
Chapter 3

Market Clearing

In the framework of the thesis, a Market-Based Control strategy is used to manage the charge. With this strategy, price is the signal to control the different agents. The price, in an MBC, is determined according to the microeconomic theory of supply and demand. Moreover, in this approach, vehicles are integrated in the electricity market. They respond to the market price according to the bids they placed.

The price can be determined in two ways.

- A simple market clearing determines the market equilibrium price.
- With Optimal Power Flow (OPF), where the different Nodal prices are used as a control signal in the different nodes where the cars are connected.

The market clearing and OPF formulations are given in this chapter.

3.1 Electricity Market

In the electricity market, the market operator determines the price by matching the supply and demand bids. The goal is to maximize social welfare. According to the microeconomic theory, that happens when the consumer and producer surplus are maximized.

In the thesis, the PEV charging scheme is integrated into two different types of dispatch. First, the equilibrium of the market is found by solving a copper plate market clearing. The only constraint is the equilibrium between generation and demand. Second, the integration into an OPF framework is used to include network constraints.
3.1.1 Copper plate market clearing

The market clearing problem for each time step is defined as follow:

$$\max \sum_{L_{t}} L_{i} \times P_{L_{i}} - \sum_{G_{t}} G_{i} \times P_{G_{i}}$$ (3.1)

subject to

- $$P_{L_{t},min} \leq P_{L_{t}} \leq P_{L_{t},max}$$
- $$P_{G_{t},min} \leq P_{G_{i}} \leq P_{G_{t},max}$$
- $$\sum_{L_{t}} P_{L_{i}} = \sum_{G_{t}} P_{G_{i}}$$

Where:

- $$P_{G_{i}}$$ is the output of the generator $$G_{i}$$ at time $$t$$
- $$P_{L_{i}}$$ is the demand of the load $$L_{i}$$ at time $$t$$
- $$p_{G_{i}}$$ is the bid price of the Generator $$G_{i}$$ at time $$t$$
- $$p_{L_{i}}$$ is the bid price of the Load $$L_{i}$$ at time $$t$$

The clearing price of the market is defined by the Lagrange multiplier of the equilibrium constraint between generation and demand.

3.1.2 Optimal Power Flow

In the previous case, apart from the bounds on generators and loads, the only constraint is the equality constraint of the power balance between producers and consumers.

In the case of an OPF, the characteristics of the network are taken into account. Loads and generators are connected to different nodes and the topology and characteristics of the grid are taken into account. The OPF formulation used in this work is a DC-OPF using Power Transfer Distribution Factors. The inequality constraints to avoid asset overloads are given by:

$$\left| \sum_{n} PTDF_{n,m} \cdot (P_{G_{i} \in \Omega_{n}} - P_{L_{j} \in \Omega_{n}}) \right| \leq P_{n,max} \forall n$$ (3.2)

Where:

- $$P_{G_{i} \in \Omega_{n}}$$ is the power injected by the generator $$G_{i}$$ connected to the node $$n$$.
- $$P_{L_{j} \in \Omega_{n}}$$ in the power consumed by the load $$L_{j}$$ connected to the node $$n$$. 
3.1. **ELECTRICITY MARKET**

- $P_{l_{m_{\text{max}}}}$ is the power flow through line $l_m$.
- $\text{PTDF}_{n,l_{m}}$ are the factors used in DC-OPF approximations [9] to modelize the changes on the flow in a line when an specific amount of power is injected in one node.

The optimization problem of the OPF is set by the combination of 3.1 and 3.2

An example of the use of an OPF in the context of smart charging is found in [10]. In an OPF the different nodal prices are calculated on the Lagrange multipliers of the power flow and supply-demand balance constraints.
Chapter 4

Reinforcement Learning: Q-Learning

Reinforcement learning is one of the three different Machine learning algorithms. The other two are supervised learning and unsupervised learning [11]. In supervised learning, also known as predictive learning, the agent learns from a training set and generates a mapping from inputs to outputs. In unsupervised learning the agent learns from unlabeled data. The goal of the agent is to find a pattern in the data and classify it.

In reinforcement learning, the learning is based on the reward signals obtained after performing an action. Those rewards are used to build a policy that allows the agents to choose the best future option. The maximization of the reward is the goal of the algorithm.

Q-learning is a Reinforcement Learning algorithm, first developed by Watkins [12]. In this method the agents can perform different actions while being in one of the different available states. The reward is assigned to the particular state-action pair. Those values are the Q-values, and the goal is to find the strategies to maximize them.

The aim of this chapter is to describe the main Q-learning algorithm. The parameters and variables are presented, as well as the iteration process. An important step for the learning is the process of choosing the action to be performed. For that purpose we use the $\epsilon$-Greedy policy. Finally, the implementation of the algorithm, focusing in the particular case of this thesis is presented. However, the setup of every parameter will be done in Chapter 5.
4.1 Q-learning algorithm

The learning process uses an iterative algorithm to update the Q-value, that is the reward of a given state-action pair. Initially, Q-values corresponding to each state-action pair are chosen by the designer. Those values change when the agent performs an action. A description of the iterative process can be found in [13].

As presented there, the update is done according to this equation:

\[
Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha_t (s_t, a_t) \times \left[ R_{t+1} + \gamma \times \max_a Q_t(s_{t+1}, a) - Q_t(s_t, a_t) \right] (4.1)
\]

where:

- \( Q_t(s_t, a_t) \) is the old Q-value, i.e. the Q-value of the iteration \( t \), \( Q_{t+1}(s_t, a_t) \) is the updated Q-value,
- \( a_t \) is the particular action \( a \in A \), performed in the iteration \( t \),
- \( s_t \) is the state \( s \in S \) in which the agent is during the iteration \( t \),
- \( R_{t+1} \) is the reward obtained, after performing the action \( a \),
- the learning rate \( \alpha \) is a value between 0 and 1, and it determines the weight of the new information acquired versus the old information,
- the term \( \max_a Q_t(s_{t+1}, a) \) is an estimation of future values,
- \( \gamma \) is the discount factor. The discount factor is a weighting coefficient of future or actual rewards.

4.2 The \( \epsilon \)-Greedy policy

The application of the Q-learning algorithm allows the agent to know which are the most profitable actions. However, no rule about the decision process is given in the algorithm. Hence, the use of a policy for the decision is required.

The agent can exploit the available information using a Greedy policy. With that policy the agent always choose the action that maximize the reward. However, with this behavior, the agent will try the more promising actions while the rest remain unexplored. Thus, the question between Exploration or Exploitation arises. The decision between those two is based on the certainty of being in the best possible state or the interest to explore...
new actions. In [14] different strategies are given. The conclusion is that the trade-off is very problem dependent.

The use of a $\epsilon$-Greedy policy, as in [6], is a good strategy to trade-off these options. When using this policy, the agent will perform the action that maximizes the reward with probability $1-\epsilon$ and, with a probability $\epsilon$, a random available action $a \in A$ will be chosen.
Chapter 5

The Market-based control model

Chapter 2 is concerned with the use of a MAS to describe a system with a high number of active elements and the combination with MBC to manage such a complex system. It is shown that both can lead to a good coordination and control of the agents of the system.

Here we describe how these concepts are applied to the control of vehicle charging. An important feature of the vehicles is that they are considered as flexible loads. Flexible loads are those whose demand can be adapted and is shiftable. In the case of a PEV fleet, vehicles can purchase the required energy at any time prior to the departure time. That allows the use of smart-charging strategies to shape the demand of the vehicles.

Based on the concepts of MAS and MBC, the different agents, and the control scheme developed in this work are presented in this chapter. In the Figure 5.1 the multi-agent hierarchical structure is presented.

As shown in that figure, the system is modeled with a logical tree layout with three levels. The PEV agents are on the lower level. The intermediate level corresponds to the aggregator agent, which acts as a middleman in terms of communication between PEVs and the higher level. The agent at the top level corresponds to the electricity market.

The communication between agents is presented in the Figure 5.1 as well. It is based on the bids that the vehicles place according to their requirements. Those bids are aggregated at the intermediate level, and finally communicated to the market level.

The MBC is a control strategy based on the micro-economic theory of
demand-supply, where the different agents of the system compete for the resources according to this. The equilibrium price of the market is the control signal. Here, this price is the clearing price of the day-ahead electricity market.

5.1 PEV agents

As in a decentralized scheme, the vehicles in this approach try to minimize their charging cost. They participate in the day-ahead market by placing bids. As they have to optimize their cost, they need to learn how to place optimal bids to compete with the rest of the agents and, at the same time, take profit of their condition of flexible loads.

Different data is known about the vehicles, describing both the battery and the driving patterns of the fleet.

5.1.1 Definition of the Demand Bids

The bids should express the willingness or need of the user to buy a certain amount of energy. Following the idea introduced in [4], a piecewise linear function is used to model the demand curve of the vehicle.
5.1. PEV AGENTS

The function, as presented in the paper, has two different intervals. The first interval is an inelastic demand interval. The second interval of the function presents a linear relation, characterized by the so-called corner priority of the PEV. The corner priority is the urgency of the PEV to charge, and defines the slope of the function in the interval.

In the approach developed in the thesis the demand curve of the individual vehicles is a piecewise linear function. In that case, the function is defined in three different intervals, as follows:

\[
\text{demand}(P) = \begin{cases} 
3000 : & 0 \leq P \leq P_{\text{min}} \\
p_{\text{state}} : & P_{\text{min}} \leq P \leq P_{\text{int}} \\
p_{\text{int}} : & P_{\text{int}} \leq P \leq P_{\text{max}} 
\end{cases}
\] (5.1)

The values of the power and the price that defines the demand are based on the state variables of the vehicle in that particular time step.

The definition of the price \(p_{\text{state}}\) is similar to the corner priority in [4]. The value expresses the urgency of the vehicle to buy the energy. This price is defined as follows:

\[
p_{\text{state}} = \frac{E_{\text{required}}}{i2d \times P_{\text{connect}}}
\] (5.2)

In that case, \(E_{\text{required}}\) is the energy the vehicle needs to charge before the next trip. That is, the difference between the energy level that the battery should reach before departure and the energy that the vehicle has already charged.

The value \(P_{\text{max}}\) corresponds to the maximum power the battery can be connected at in order to charge it completely during the time that the vehicle is charging. This value is limited by \(P_{\text{connect}}\), as it is the maximum power that the battery can be connected at.

The value \(P_{\text{min}}\) is defined as the minimum power that a PEV should charge at a given time step in order to reach the necessary SOC before departure. This value is typically zero for most vehicles most of the time.

Finally, the pair \((P_{\text{int}}, p_{\text{int}})\) is defined as the intermediate point between \((P_{\text{min}}, p_{\text{state}})\) and \((P_{\text{max}}, 0)\). That way, the agents only have to calculate three values according to their state, \(P_{\text{min}}, p_{\text{state}}\) and \(P_{\text{max}}\), and the rest are derived from them.
The demand curve defined by that function is represented on the Figure 5.2.

There are different demand profiles depending on the status of the vehicle. Those different types are defined as follow:

No charge: There are two possible reasons why a vehicle cannot further charge the battery. Vehicles not connected or vehicles with the battery fully charged, i.e, the energy at the beginning of the time step is equal to the maximum capacity of the battery ($C_{batt}$). In that case $P_{min} = P_{max} = 0$

Urgent: A vehicle needs to charge urgently when it has not enough time to reach the required state of charge before the departure in case the vehicle has to leave one time step earlier than expected. In that case $P_{min} > 0$

Not Urgent: When the vehicle has enough time to charge even if forgoes charging at the current time step, it is consider a non-urgent PEV. In that case $P_{min} = 0$

No Required: Finally, the vehicles connected and whose state of charge is higher than that required before the departure are vehicles that do not need to charge, but they can charge because the energy level of the battery is lower than the maximum capacity of the battery. Those cars will charge in case the clearing price of the market is low.
The Table 5.1 summarizes the previous paragraphs:

<table>
<thead>
<tr>
<th>Action</th>
<th>( t2d )</th>
<th>( t2d - t_{\text{step}} )</th>
<th>( \text{SOC}_{\text{req}} )</th>
<th>( E_{\text{init}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Charge</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>( C_{\text{batt}} )</td>
</tr>
<tr>
<td>Urgent</td>
<td>( &gt; 0 )</td>
<td>( \leq \text{SOC}_{\text{req}} )</td>
<td>( &gt; \text{SOC}_{\text{real}} )</td>
<td>( &lt; C_{\text{batt}} )</td>
</tr>
<tr>
<td>Not urgent</td>
<td>( &gt; 0 )</td>
<td>( &gt; \text{SOC}_{\text{req}} )</td>
<td>( &gt; \text{SOC}_{\text{real}} )</td>
<td>( &lt; C_{\text{batt}} )</td>
</tr>
<tr>
<td>No required</td>
<td>( &gt; 0 )</td>
<td>-</td>
<td>( &lt; \text{SOC}_{\text{real}} )</td>
<td>( &lt; C_{\text{batt}} )</td>
</tr>
</tbody>
</table>

Therefore, the bidding profile of the different vehicles is defined according to the status of the vehicle in that precise time step. PEVs behave as intelligent agents, taking different actions as explained in the next subsection. The definition of the bid can be modified by the vehicle by taking different actions.

### 5.1.2 Actions

By choosing between the different actions vehicles are able to adapt their bids to the market prices. Two different actions can be chosen by the agents. The actions modify directly the price of the bid, i.e. the value of \( p_{\text{state}} \). The new definition, including the actions is:

\[
p_{\text{state}} = D_p \times \frac{E_{\text{required}}}{t2d \times P_{\text{connect}}} + C_p
\]

To modify the price, the agent can increase the price by a constant value \( C_p \). The vehicles can set their interest in charging for that day increasing the price of the bid.

Moreover, the different values of the action \( D_p \) increase the bid price proportionally to the price defined by the status variables of the PEV. This factor modifies the price according to the urgency, defined according the status variables of the vehicle.

The changes that the actions produce should allow the optimization of charging costs. The agent has to learn how to place the optimal bid by modifying \( D_p \) and \( C_p \). The use of a machine learning algorithm intends to lead the PEV agents to an optimal bidding policy.
5.2 Aggregator agent

The integration of the electric vehicle into the electric power system poses different challenges. The aggregator agent is envisioned to be an important element for this integration. This agent should play a key role as a link between the market and the vehicles [15].

Different frameworks include the role of an aggregator agent. For instance [16] and [17]. In the framework of [17] the main function of such an agent is to group the electric vehicles, and try to satisfy the requirements of the fleet. This agent should lead to an efficient combination of centralized and local control. It is supposed to favour the integration of the different services that the vehicles offer.

However, in this work, the aggregator’s role is similar to the role of the concentrator agent in [2]. As shown in the Figure 5.1 in this approach the aggregator acts as intermediate agent between the fleet and the market. It is a bidirectional communication agent. Bids are received, concentrated and sent to the market. The aggregator acts on behalf of the vehicles in the market. On the other way, the clearing price of the market is communicated to the vehicles through this agent. With this information vehicles will charge or not according to the market equilibrium and they requirements at that moment.

In addition, such an agent could have other functions. For instance, in the case of different aggregator agents at different hierarchical levels (e.g. at different voltage levels), those agents can take care of network constraints.

5.2.1 Aggregating process

The aggregated demand is obtained by the horizontal summation of the different demand curves of each vehicle. As in [10], the aggregated function is a piecewise linear function as well. It has as many intervals as different prices are found in the individual demand curves.

The bid aggregation process is described in the Figure 5.3.
5.2. AGGREGATOR AGENT

Figure 5.3: Individual and aggregated demand curves
CHAPTER 5. THE MARKET-BASED CONTROL MODEL
Chapter 6

Implementation

The implementation of the simulations is presented in this chapter. First, an introduction to the general scheme of the simulation with the different steps is given. The simulation is implemented with Matlab, in the Appendix C describes the code developed for the simulation. Moreover, a description of the Q-learning algorithm used in this work is given in this chapter. The general Q-learning was presented in the Chapter 4 here the particular application in this thesis is presented.

6.1 General implementation

The four study cases are presented in the next chapter. Irrespective of the particular considerations, the general implementation of the simulation has the following steps:

1. At the beginning of the day, the actions that each vehicle will perform are determined according to the $\epsilon$-Greedy policy.

2. The bids defining the willingness of the vehicle are defined at the first level, and aggregated at the intermediate level.

3. Then the market clearing, with the integration of the vehicle bids with the rest of the demand and the supply bids is performed.

4. After the equilibrium is determined, the vehicles update the state of charge according to the amount of energy they bought. At the same time, the reward of the action is updated.

5. At the end of the day the Q matrix is updated and a new action is chosen for the next day.

This general implementation is presented in the flow diagram of the Figure 6.1.
6.2 Implementation of the Q-learning

In the implementation of the Q-learning in the thesis is important to remark that no differentiation between the states that vehicles take at each iteration is done. The reason is that the different status are already taken into account in the parametrization of the bid.

This has two important consequences. First, the reward is always assigned to the action that the vehicle performs in that precise moment. Thus, the values can be stored in a matrix with the number of cars as rows and the number of possible actions as columns.

\[
Q_{\text{matrix}} = \begin{pmatrix}
Q_{1,1} & \cdots & Q_{1,nactions} \\
\vdots & \ddots & \vdots \\
Q_{ncars,1} & \cdots & Q_{ncars,nactions}
\end{pmatrix}
\] (6.1)

The other important consequence is that future rewards are not taken into account. That means that the Q values are associated to the action performed and the term \(\max_a Q_t(s_{t+1}, a)\) is not necessary in the algorithm.

Thus, the update algorithm will be:

\[
Q_{t+1}(a_t) = Q_t(a_t) + \alpha_t(a_t) \times [R_{t+1} - Q_t(a_t)]
\] (6.2)
The Reward is calculated according to the following equation:

$$ R = - \sum_{t=1}^{T} (\text{price}(t) \times \text{Power}(t)) + \text{Penalty} \times (E_0 - E_{24}) $$  \hspace{1cm} (6.3)

On the one hand, the first term is the sum of the energy purchased by the vehicle times the price at every time step during the day. The price is the clearing price of the market or the nodal price of the node where the vehicle is connected to.

On the other hand, the penalty term, takes into account the state of charge profile of the cars during the whole day. Cars that do not buy enough energy to cover their daily need are penalized. This term aims at capturing the future cost of the energy. $E_0$ is the energy stored in the battery at the beginning of the day and $E_{24}$ is the energy stored in the battery at the end of the day.

In this implementation, the values for the learning rate, $\alpha$, change as the iterative process advances. At the beginning, while most of the actions are unexplored, a high $\alpha$ allows the agent to get more information from the rewards. The learning rate decreases as the optimal actions are known.

The different actions that the cars perform in every iteration are chosen according to an $\epsilon$-Greedy policy. One important feature of the Q-learning is the convergence condition. The algorithm finds the optimal policy only if all the actions have been taken a sufficient number of times. Thus, the main reason to apply the $\epsilon$-Greedy policy is to make sure that all the actions are performed.
Chapter 7

Results

The scheme studied in the thesis is finally tested under different scenarios. In this chapter the four different cases studies are presented, in the following order:

1. Base case
2. Integration into an OPF market clearing
3. Uncertainty in the behavior of the PEV fleet.
4. Uncertainty in demand and supply.

Each case is presented in the corresponding section, as well as the different goals of each simulation. The results are evaluated and discussed in the corresponding part. First, the base case, where the multi-agent approach is tested, is presented. In the second case the grid constraint are included in the formulation, integrating the approach into an OPF. The other two cases introduce some uncertainties. The goal is to test the approach under different patterns of the fleet behavior (third case) and a real demand-supply, as in the fourth case.

7.1 Base case

The purpose of this first case is to test the approach and validate the control scheme. At the same time, the setup of the parameters and the study of the convergence of the Q-learning algorithm are the two main focuses of this first simulation.

In this case, the same demand and supply data is used in every iteration. Moreover, the demand of the reference load is inflexible. Regarding
the demand of the vehicles, the energy requirements and the driving patterns are invariant during the whole simulation. The market equilibrium is determined as a copper plate market clearing.

### 7.1.1 Results and comments

In Figure 7.1, the evolution of the system load, the market prices and the charging profile of the fleet are presented. An scenario without vehicles is used as a reference of the system load and prices, and two scenarios with the integration of vehicles are compared.

First case corresponds to an inflexible demand, while the other case corresponds to the use of a bidding strategy.

![Figure 7.1: Base case: System load, market prices and charging profile of the fleet](image)

In the Figure 7.1 (a), it is shown that with the bidding strategy the charging during the peaks is avoided. The Figure 7.1 (c) shows the demand
of the vehicles and, in the case of the bidding strategy it has been mainly displaced to the valley hours, where the prices are lower. In the case of the inflexible demand, the load is distributed along the day, and it has two peaks almost simultaneous to the demand peaks. With the bidding strategy the demand is mainly displaced to the night hours, but still a peak appears between the two daily peaks.

Moreover, the Figure 7.2 compares the average cost of the energy purchased by the vehicles. This price is calculated according to:

$$\text{Price} = \frac{\sum_{t=1}^{n} \sum_{i=1}^{24} \text{price}(t) \times \text{Power}(t)}{\sum_{t=1}^{n} \sum_{i=1}^{24} \text{Power}(t)}$$

(7.1)

As shown here, the prices are lower when the strategy is applied. Moreover, the prices are lower at the end, when the agents are supposed to exploit their knowledge about placing the optimal bids.

In what refers to the convergence, as the number of iterations increases, the Figure 7.1 shows that some patterns are established for the load, the
price and the shape of the demand of the fleet. In the Figure 7.3 the evolution of the total reward of the fleet each iteration is presented. According to the definition of the reward for each vehicle, it is the cost of the energy purchased including the penalty term. The value represented is the total cost of the system in each iteration, taking into account also the estimated future cost. The system tends to find the optimal policy, as the value of the system cost tends to be lower.

![Figure 7.3: Base case: Total reward for the whole fleet](image)

The Figure 7.3 shows the different actions that an agent has performed during the simulation. During the first iterations the agent is trying all the different actions and after this period it seems that the agent finds optimal actions to exploit.

In the Appendix B a more detailed study of the convergence, taking into account the parameters of the Q-learning is done. The results with different $\epsilon$-greedy policies and the learning rates are presented there.

### 7.2 Integration into an OPF market clearing

The second case study corresponds to the integration of the control scheme into an OPF. Since in this case the topology of the network is taken into account, the number of aggregator agents corresponds to the number of dif-
7.2. INTEGRATION INTO AN OPF MARKET CLEARING

Figure 7.4: Base case: actions performed by one of the vehicles during the simulation

ferent nodes in the system. In this case, the different nodal prices are the control signals for the charging of the fleet.

As in the previous case, demand and supply are equal at every iteration and the demand of the reference load is inflexible.

7.2.1 Results and comments

In the Figure 7.5 (a), the total load of the system in the case of an inflexible demand and the bidding strategy are compared after three hundred iterations. Again, the charging during the peaks is avoided using the bidding strategy. In the Figure 7.5 (b), the charging profiles of the fleet in both cases are presented.

Moreover, the next three Figures, 7.6, 7.7 and 7.8 show the nodal prices of the three scenarios. Comparing the two scenarios with vehicles, the case of the bidding strategy leads to lower nodal prices. As the profile of the prices is narrower in the case of the bidding strategy, the result is also good in terms of grid congestion.

In this case, a further step for the control mechanism should be to differentiate the actions according to the node where each car is connected to. As the definition of the actions is a key issue to have a good control of the fleet, a more precise definition in this case should lead to a better result
CHAPTER 7. RESULTS

(b) Charging profiles

(a) Traded volumes

Figure 7.5: Integration into an OPF: System load and charging profile of the fleet

Figure 7.6: Integration into an OPF: Nodal prices in the scenario without PEVs
7.3 Uncertainty in the behavior of the vehicles

This third case adds to the base case the uncertainty in the behavior of the fleet, while demand and supply remain the same. According to the available

Figure 7.7: Integration into an OPF: Nodal prices in the scenario with inflexible demand of the PEVs

Figure 7.8: Integration into an OPF: Nodal prices in the scenario with the use of the bidding strategy

In terms of shaping the demand of the fleet.

In the Appendix B a further study of this case is done. As the prices in the OPF are more diverse, the use of different values for the penalty term in the Q-learning lead to different results in terms of the charging profile of the fleet.
data, a hundred different patterns for each vehicle’s behavior are available.

### 7.3.1 Results and comments

The Figure 7.9 shows the load of the system, the market and the charging profile of the fleet, in the three scenarios, without vehicles, with an inflexible demand and in the case of the bidding strategy.

Figure 7.9: Vehicles behavior uncertainty: System load, market prices and charging profile of the fleet

The results are similar to the base case, the demand of the fleet is mainly displaced to the night hours and the charging during the peaks is avoided. Moreover, the Figure 7.10 shows that the cost of the electricity converge to a value around 74.5 €/MWh, lower than the price in the case of an inflexible
7.3. **UNCERTAINTY IN THE BEHAVIOR OF THE VEHICLES**

demand. Finally, the Figure 7.10 shows the value of the reward of the whole fleet.

![Figure 7.10: Vehicles behavior uncertainty case: Evolution of the cost of the energy](image)

Figure 7.10: Vehicles behavior uncertainty case: Evolution of the cost of the energy

![Figure 7.11: Vehicles behavior uncertainty case: Total reward for the whole fleet](image)

Figure 7.11: Vehicles behavior uncertainty case: Total reward for the whole fleet
CHAPTER 7. RESULTS

7.4 Uncertainty in demand supply

In this case, demand and supply are based on the real data of the EEX for the years 2011 and 2012 in Germany and Austria. Moreover, the PEV penetration is assumed to be 2%. The data used in this case study is the same as in [18].

In this paper a PEV fleet aggregator is studied, which will be responsible for managing the charge and purchase the energy. The benefits of such an agent acting on behalf of the vehicles are studied. As the data used is the same, it is interesting to compare both concepts.

In the simulation of this case, three different Q-matrices are used, so there is a differentiation between weekdays, Saturday and Sunday. Three different sets of actions are used as well, according to the average prices of each of the three groups of days. The data of the year 2011 is used to train the algorithm, so the initial Q-matrix for the simulation of 2012 correspond with the values obtained from this first simulation.

7.4.1 Results and comments

The Figure 7.12 shows the load profile of the demand including the vehicles, the market prices and the energy purchased by the fleet during one week of 2012.

While the charging profile of the fleet in the case of inflexible demand is distributed during the day, in the case of the bidding strategy the profile has two peaks, being the larger one during the night. During the day the profile of this second case has a peak as well. However, the demand during the day is lower than in the case of inflexible demand, and the peaks take place mostly between the two daily peaks.

In this case, it is interesting to compare the cost of the energy with the bidding strategy and the inflexible demand of the PEVs. The Table 7.1 shows the values of the average price of the energy purchased by the fleet. With the bidding strategy, the price is 36.31% lower.

<table>
<thead>
<tr>
<th>Table 7.1: Cost comparison (€/MWh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inflexible demand</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>48.46</td>
</tr>
</tbody>
</table>

The comparison with [18], where the role of an aggregator agent for the
7.4. UNCERTAINTY IN DEMAND SUPPLY

(b) Market prices

(c) Charging profiles

Figure 7.12: System load, market prices and charging profile of the fleet for the week Junet 4th-June 10th, 2012

fleet is studied, leads to some interesting conclusions.

First, the charging cost with an aggregator agent are reduced comparing with the strategy where the vehicles can decide when to charge. Here, the average cost is 36.82 €/MWh, while with the bidding strategy of the aggregator agent, the cost is 34.55 €/MWh.

Comparing the charging profile of the fleet, the results also show that the aggregator agent leads to a better profile, as the charging is mostly moved to the valley hour of the night. From a cost-minimization perspective, having an aggregator agent optimize the charging for the fleet might be a better
solution than letting each vehicle optimize for itself.

As presented in the Appendix A, the definition of the actions, mainly the $C_p$ value, has to be done in a tight range of prices. With the demand-supply uncertainty those values are more difficult to determine, as there is more diversity in the price values. A more precise setting of this actions is likely to lead to a better control of the charge, reducing the cost and increasing the peak of the charge during the night and reducing the charge during the day.
Chapter 8

Conclusions

In this thesis a smart-charging strategy for the PEV fleet is introduced. The scheme is based on the multi-agent theory. A market-based control is used to manage the fleet where each vehicle tries to minimize its costs of charging.

The approach is tested in four different case studies. The results show that the proposed control scheme is able to shape the demand of the fleet, displacing the charging mainly to the night hours, when the price of the energy is lower. Moreover, the charging during the demand peaks is avoided.

In terms of convergence, the Q-learning algorithm seems to converge and allows the vehicles to find the optimal bidding strategy. It is shown both in terms of reducing the cost of the energy and behavior of the vehicles when defining their bids.

The control scheme is proven to lead to good results even with the introduction of some uncertainties. For instance, in the case with uncertainty in the behavior of the fleet, with varying energy requirements and driving patterns, the scheme has a good performance in terms of convergence, shaping the demand and reducing the cost of the energy purchased by the fleet.

In addition, when integrated into an OPF, the results show that a scheme with multiple aggregator agents can be managed. The nodal prices are lower with the bidding strategy, and the profile of the prices is narrower in the case of the bidding strategy, so the result is also good in terms of grid congestion.

Finally, when the uncertainty in the demand-supply is introduced, using the real data from 2012, the response of the control scheme is satisfactory as well. The cost of the energy is lower with the bidding strategy and the demand is displaced to the night hours. However, this case can be compared with [18] as the same data is used. In this paper the bidding strategy
is performed by an aggregator agent. The results of both works show that the figure of an aggregator agent, that take cares of the fleet requirements and tries to optimize the cost, leads to a lower costs and a better charging profile of the fleet.

8.1 Outlook

A better definition of the actions that the vehicles can perform to set the price of their bids should be the focus of the future work.

In the case of the OPF, different sets of actions, according to the node where the vehicle is connected should lead to more accurate results. In the case of the real demand-supply of 2011 and 2012, three different sets of actions are being used, making a distinction between weekdays, Saturdays and Sundays. A way to have a more precise definition of the actions, so the vehicles can state their requirements in a more optimal way, could be to define different sets of actions according to the season of the year, or the different months.

As the comparison with [18] shows, the figure of an aggregator leads to better results. A more active role of the aggregator agent introduced in this work could be an interesting way to achieve better results in terms of minimizing the cost of charging.
Bibliography


Appendix A

Definition of the actions

An important aspect addressed in this thesis is the definition of the demand bids of the PEVs agents. On the one hand, the bids should express the willingness of the vehicles to purchase the energy. On the other hand, as the vehicles try to optimize their cost, they should learn how to place the optimal bids. For that reason, in the definition of the bids two parameters allows them to modify their bids as shown in the equation.

\[
p_{state} = D_p \times \frac{E_{required}}{t2d \times P_{connect}} + C_p \tag{A.1}
\]

The definition of this parameters, here called actions, is already done in the Chapter 5. \( C_p \) is the parameter that sets the reference price, while \( D_p \) takes into account the requirements of the vehicle. In this appendix, a more detailed definition of the values of the two parameters is given.

A.1 Definition of \( D_p \)

In the Equation A.1 \( D_p \) multiplies the term based on the state variables of the vehicles. That way, the purpose of this parameter is to measure the importance of the urgency term.

Three different values for \( D_p \) are defined
A.2 Definition of $C_p$

The second parameter, $C_p$, modifies the price of the bid a constant value. This parameter is defined according to the prices of the market. In general, the definition is done according to the clearing price of the market.

The definition of $C_p$ for the base case and the case with uncertainty in the behavior of the vehicles, where demand and supply are always the same, is done according to the prices shown in the Figure A.1.

![Figure A.1: Clearing price of the market without vehicles](image)

According to the Figure A.1 the values of $C_p$ should be between 60 and 85. A total of twelve values are defined, ten between that range and the other two are values below and above the range. However, the results in the Figure A.2 shows the results with two different definitions for $C_p$. If the values are defined by regular steps within the range determined by the minimum and the maximum, the Figure A.2 shows that the results in terms of shaping the demand are similar to the case with no strategy. Therefore, a more accurate definition is required.

When the values of $C_p$ are defined in regular steps within a range defined by the clearing prices during the valley hours the Figure A.2 shows better results in terms of shaping the demand. In this case, twelve values are defined as well. Ten of them are between this interval and the other two, as done before, are below and above the range defined by the clearing price of the day. The charging profile of the vehicles is displaced to the night hours.

The other two cases, the integration into an OPF and the case with the uncertainty in demand and supply, require a more complex definition. In this case a study of the distribution of the minimum prices is done before defining the values of $C_p$. The Figure A.3 shows the distribution of the minimum prices for the week days of 2012. According to this results the values
A.2. DEFINITION OF $C_p$

Figure A.2: Base case: System load, market prices and charging profile of the fleet with two different sets of actions of $C_p$ are defined.
Figure A.3: (a) Distribution of the minimum clearing prices of the market without vehicles in 2012 (b) Detail of (a)
Appendix B

Q-learning parameters

There are several parameters that are critical to the simulation results. In what refers to the implemented Q-learning algorithm three parameters have to be set. The learning rate $\alpha$, the $\epsilon$ of the $\epsilon$-Greedy policy and the value of the penalty.

B.1 Learning rate and $\epsilon$-Greedy policy

The goal of the definition of $\alpha$ and $\epsilon$ is to control the way the agent learns. On the one hand, $\epsilon$ controls the trade-off between exploitation and exploration. On the other hand, $\alpha$ is the weight coefficient between the information learned from the reward and the value that the agent already knows.

The values of this parameters decrease as the iterations advance, according to the idea of exploring and having a higher learning at the beginning and exploiting the knowledge when the different actions are already explored.

We use the base case scenario to test different approaches. The following figures show the results of the use of different $\epsilon$-Greedy policy and learning rate values.

First, in Figure B.1, it is shown that similar results are obtained regardless the values of $\alpha$ and $\epsilon$. A more interesting result is the one presented in Figure B.2. As the $\epsilon$ value is lower since the beginning, the exploitation is favored in the case with constant $\epsilon$. However, the two other cases, with more exploration at the beginning lead to better values for the reward.

Finally, the Figure B.3 shows the actions performed by an agent with different values of $\alpha$ and $\epsilon$. Regardless the values, the agent has a first period where it explores all the different actions, then progressively it finds actions
APPENDIX B. Q-LEARNING PARAMETERS

Figure B.1: Base case: System load, market prices and charging profile of the fleet with different $\epsilon$-Greedy policies and learning rate values

Figure B.2: Base case: Reward values with different $\epsilon$-Greedy policies and learning rate values
B.2 PENALTY TERM

The other parameter that can be modified is penalty term of the reward. This value is used to take into account the future cost of the energy for the vehicles. In the case of the OPF the different nodal prices present a broader range of values. Thus, three different simulations are run using the OPF market clearing approach.

As shown in Figure B.4 (b), the peak on the charging profile during the night is higher. At the same time, the profile during the day hours with the higher penalty term is below the one corresponding to the lower penalty. With a higher value for the penalty, the agents tend to choose the higher actions among the possible, while a lower penalty term the agents tend to choose the lower actions. This behavior is shown in the Figure B.5 (a), where the actions that one of the agents has performed during the day are presented with the three different penalty values.

However, the different values of the penalty term do not have an important influence on the convergence. The Figure B.5 (b) shows the values of the reward in the three different cases.
Figure B.4: Integration into an OPF: System load and charging profile of the fleet with different penalty policies
B.2. PENALTY TERM

Figure B.5: Integration into an OPF: Actions and Reward with different penalty policies
APPENDIX B. Q-LEARNING PARAMETERS
Appendix C

Matlab code

In this appendix a list with the different programs used in the simulations of each case is given.

- aggregator1Q.m: This program generates the bids from the data of each vehicle and also the action the vehicle is performing.
- aggregator2Q.m: This program acts as an aggregator of the demand regardless the nodes where the vehicles are connected.
- aggregator2QOPF.m: This program acts as an aggregator of the demand, taking into account the different nodes of the system.
- aggregator3.m: this program allocates, according to the bid of the vehicles and the clearing price of the system, the quantity that each vehicle can charge.
- aggregator3OPF.m: this program allocates, according with the bid of the vehicles and the nodal prices of the system, the quantity that each vehicle can charge.
- mktclr.m: Calling for the optimization routine of the market as a copper plate market clearing.
- ptdf_market_clearing.m: Calling for the optimization routine of the market as an OPF.

Four different cases of study are presented in this thesis. The Table C.1 shows the programs used to call the simulation routine and the simulation routine of each case. The Table C.2 shows the different programs used in each simulation case.
Table C.1: Code

<table>
<thead>
<tr>
<th>Study case</th>
<th>Simulation code</th>
<th>Calling for the optimization code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base case</td>
<td>Q_gen.m</td>
<td>Q_generator.m</td>
</tr>
<tr>
<td>Integration into an OPF</td>
<td>Q_gen_OPF.m</td>
<td>Q_generator_OPF.m</td>
</tr>
<tr>
<td>Vehicles behavior uncertainty</td>
<td>Q_gen_unc.m</td>
<td>Q_generator_unc.m</td>
</tr>
<tr>
<td>Demand-Supply Uncertainty</td>
<td>Q_gen_anual.m</td>
<td>Q_generator_anual.m</td>
</tr>
</tbody>
</table>

Table C.2: code

<table>
<thead>
<tr>
<th>Study case</th>
<th>Bid generator</th>
<th>Bid aggregator</th>
<th>Market clearing</th>
<th>Power allocation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base case</td>
<td>aggregator1Q.m</td>
<td>aggregator2Q.m</td>
<td>mktclr.m</td>
<td>aggregator3.m</td>
</tr>
<tr>
<td>Integration into an OPF</td>
<td>aggregator1Q.m</td>
<td>aggregator2QOPF.m</td>
<td>ptdf_market.clearing.m</td>
<td>aggregator3OPF.m</td>
</tr>
<tr>
<td>Vehicles behavior uncertainty</td>
<td>aggregator1Q.m</td>
<td>aggregator2Q.m</td>
<td>mktclr.m</td>
<td>aggregator3.m</td>
</tr>
<tr>
<td>Demand-Supply Uncertainty</td>
<td>aggregator1Q.m</td>
<td>aggregator2Q.m</td>
<td>mktclr.m</td>
<td>aggregator3.m</td>
</tr>
</tbody>
</table>