# Learning Strategies for Power Trading in Smart Grids

Thesis submitted in partial fulfillment
of the requirements for the degree of

*Master of Science*
*in*
*Computer Science and Engineering*
*by Research*

by

Susobhan Ghosh
201503005
susobhan.ghosh@research.iiit.ac.in

International Institute of Information Technology
(Deemed to be University)
Hyderabad - 500 032, INDIA
July 2019

International Institute of Information Technology
(Deemed to be University)
Hyderabad, India

# CERTIFICATE

It is certified that the work contained in this thesis, titled "**Learning Strategies for Power Trading in Smart Grids**" by **Susobhan Ghosh**, has been carried out under my supervision and is not submitted elsewhere for a degree.

| | |
|---|---|
| Date | Adviser: Prof. Praveen Paruchuri |
| Date | Co-Adviser: Prof. Sujit Gujar |

To my parents and friends

# Acknowledgments

Firstly, I would like to thank my advisors, Prof. Praveen Paruchuri and Prof. Sujit Gujar for their continued guidance and support for the past three years. I will always cherish the fun conversations I had with Prof. Sujit, whenever I went to meet him. He has an ability to figure out and focus on the most important aspect of any work, which I try to emulate everyday. I will always value Prof. Praveen's judgement to ask the right questions in any matter. Their technical advice steered me in the research direction which this thesis entails. Moreover, their belief in my ability and their constant encouragement, both at the professional and personal level, motivated me to see this work through. I'll forever be grateful to them for bringing out the researcher in me. It truly has been an honor working with them.

Second, I really appreciate Dr. Easwar Subramanian and Dr. Sanjay Bhat for their kind and valuable inputs throughout. It was interesting to discuss ideas with them during our meetings. Also, I want to particularly thank Dr. Easwar for putting up with me, and it was fun brainstorming ideas and figuring out implementations. I value his honest and straightforward opinion on any matter - professional or personal. I would also like to thank my lab senior, Tarun Gupta for his suggestions and advice, whenever I needed it. Moreover, I would like to thank IIIT for providing me with this opportunity. I would like to thank the staff for providing lab spaces, housing, food and other necessary things during my stay on campus. I am grateful to all the professors who have taught me. I would also like to thank the IIIT administrators for helping me with the administrative procedures, TA stipends, RA stipends, conference travel documentations and support, and official documentations whenever I needed it.

I wish to thank my parents, Jhuma and Susanta Ghosh. Their continued trust and confidence in me made me believe that I can achieve anything I pursue. Words cannot describe my gratitude towards them. I would also specially thank my friends here at IIIT - Dipankar, Akshat, Prakhar, Sriharsh and Deepanshu. I've enjoyed many nights brainstorming ideas, doing assignments, and having fun conversations, over the last four years, which made this journey memorable and fun. Thank you for encouraging me all the time, and taking care of me and helping me whenever I needed it. Dipankar, thank you for being there in every team - from courses to competitions. Sriharsh, thank you for giving me company in all of the gaming nights. Also, I would like to thank my friends - Soumajeet, Mehak, Jayant, and Subhasmita for their continued belief and support, and always hearing me out. Lastly, my final year of research wouldn't have been possible without the company of Moin and Sneha. Thank you for supporting me, hearing me out at times, and helping me out whenever I needed it.

# Abstract

A smart grid is an efficient and sustainable energy system that integrates diverse generation entities, distributed storage capacity, and smart appliances and buildings. Smart grids also indirectly facilitate end-user involvement in grid stability. The brokers, supplying electricity to end-users, represent large population of customers in the electricity markets, and incentivization of such brokers to reduce their supply-demand imbalance can lead to utilization of sustainable but intermittent sources of energy like solar and wind energy. Overall, this will reduce grid operationalization costs by reducing peak demand, and reduce broker costs, and in-turn reduce energy costs for the end-users. Most of the electricity markets consist of Periodic Double Auctions (PDAs), where trillions of dollars worth electricity is traded. Any broker participating in these PDAs has to plan for bids in the current auction as well as for the future auctions, which highlights the necessity of good bidding strategies. Since the complexity and dimensionality of the actions taken by such brokers is huge, it opens avenues for deployment of autonomous energy brokers. Smart grids also bring new kind of participants in the energy markets, whose effect on the grid can only be determined through high fidelity simulations. Power TAC offers one such simulation platform using real-world weather data and complex state-of-the-art customer models. In Power TAC, autonomous energy brokers compete to make profits across tariff, wholesale and balancing markets while maintaining the stability of the grid.

We make the following contributions to the areas of smart grid, autonomous agents, game theory and machine learning applications: (1) We do a Nash Equilibrium analysis of single unit single-shot double auctions with a certain clearing price and payment rule, which we refer to as ACPR, and find it intractable to analyze as number of participating agents increase. We further derive the best response for a bidder with complete information in a double auction with ACPR. (2) Leveraging the theory developed for single-shot double auction, we proceed by modeling the PDA of Power TAC wholesale market as a Markov Decision Process (MDP), and solve it using dynamic programming. Based on that, we propose a novel bidding strategy, namely MDPLCPBS. We empirically show that MDPLCPBS follows the equilibrium strategy for double auctions that we previously analyze. In addition, we benchmark our strategy against the baseline and the state-of-the art bidding strategies for the Power TAC wholesale market PDAs, and show that MDPLCPBS outperforms most of them consistently. (3) We design a electricity usage predictor, which uses customer usage patterns and weather data, using neural networks. This is the first usage predictor which utilizes weather data in the Power TAC scenario. (4) We use an MDP based formulation for the tariff market, and solve it using Q-learning to generate tariffs. The novelty lies in defining the reward functions for the MDP, solving the MDP, and the transformation

of the solution into a tariff in the tariff market. We also propose a few heuristics which convert near-optimal fixed tariffs to time-of-use tariffs aimed at mitigating transmission capacity fees. (5) We discuss the architecture of our autonomous energy broker VidyutVanika, which was the runner-up in Power TAC 2018 Finals. VidyutVanika implements our learning strategies for the tariff and wholesale market, along with a few heuristic ideas. We have also released VidyutVanika's binary for offline testing and research purposes. (6) We do an empirical analysis of VidyutVanika's performance during Power TAC 2018 Finals, using Power TAC 2018 tournament data. We also illustrate the efficacy of its sub-modules and strategies using controlled experiments.

# Contents

# List of Figures

# List of Tables

*Chapter 1*

# Introduction

A *smart grid* is an evolved electrical system that manages electricity demand in a sustainable, reliable and economical manner, built on advanced infrastructure and tuned to facilitate the integration of all the entities involved [1]. Keeping sustainable development in mind, the world is shifting towards renewable energy resources like solar and wind energy. Smart grids are expected to improve distribution efficiency, congestion, and reliability compared to the traditional power grid, while offering sustainable, secure and clean energy supply by managing these renewable energy resources [44]. Smart grids have components such as reclosers and sensors that play a key role in reducing outages and incorporating renewable energy resources [26]. In August 2013, the Ministry of Power, Government of India, in association with India Smart Grid Forum, published a report titled "Smart Grid Vision and Roadmap for India", which lays out the roadmap for the future and implementation of smart grids in India [24]. Both US and Europe already have fully deployed smart grid technologies and markets [8] [54]. Globally, many countries are actively pushing to adopt smart grid technologies, and have laid out roadmaps to achieve the same [8] [25] [53].

In smart grids, multiple electricity generating companies (GenCos, who are sellers), and distributing agencies (referred to as *brokers*) trade electricity in the wholesale electricity markets using double auctions. In March 2018, approximately 4.7 Billion Euros worth electricity was traded in Nord Pool alone [29]. Any small improvement in the cost optimization, by deploying better bidding strategies can lead to significant improvements in the profits of the distributing agencies. Smart grids also allow customers to indirectly influence grid stability. Due to the presence of smart demand sensors in smart grids, retail brokers supplying electricity to the customers can manage their future demands and trade accordingly in the electricity markets. While the customers can't trade in such markets, brokers managing their portfolio need to aggregate their demands and trade electricity accordingly on their behalf. This provides an opportunity for brokers to make profit by managing customer costs and reducing demand-supply imbalance, which in-turn improves grid stability [15] [16] [18]. Thus, brokers who are incentivized to reduce portfolio imbalance (i.e. supply-demand imbalance) can in-turn incentivize their customers to shift their demands using tariff contracts, which will eventually reduce overall costs (for the grid as well as brokers), reduce peak demand and improve grid stability.

Such brokers need to operate profitably, while managing their portfolio imbalance and acting across multiple markets, like - (i) in the retail markets by offering retail contracts (called *tariffs*) to customers, and (ii) in the wholesale trade markets by trading electricity with generating companies and other brokers. This poses a challenging problem where a broker takes a multitude of actions across multiple markets, under real time constraints, which involve a lot of data and complex calculations, while keeping in mind the long term profitability [17] in a competitive, dynamic, and stochastic environment. This opens the avenue for smart autonomous electricity brokers in smart grids [15]. However, there are still multiple challenges in the operationalization of smart grids and autonomous brokers, like managing highly fluctuating supply-demand scenarios, engaging stakeholders with ulterior motives, and handling automation failures of participating entities.

In order to foresee such problems and examine potential solutions, Power TAC [19] provides an open source simulator platform that replicates crucial elements of a smart grid system and allows large-scale experimentation. The simulation encourages the development of autonomous broker agents that aim at making a profit by offering electricity tariffs to customers in a retail (or tariff) market, and trading energy in a competitive wholesale market with Periodic Double Auctions (PDAs), while carefully balancing their supply and demand. To this end, a Power Trading Agent Competition (Power TAC) [19] is held annually.

Machine Learning and Game Theory-based strategies are essential for such broker agents to dynamically price tariffs and predict customer usage while simultaneously placing bids in wholesale auctions. Such is the motivation behind the line of research in this thesis. The goal of this thesis is to design a learning broker with the following objectives: (i) React to competing tariffs (ii) Increase market share, i.e., subscribed customers (iii) Decrease transmission capacity costs (iv) Decrease costs of energy procurement. Each of these objectives involve developing learning algorithms which take actions depending on the feedback of previous actions. We do a game theoretic analysis of double auctions, before proposing a bidding strategy for wholesale electricity markets using PDAs and comparing it separately with state-of-the-art strategies. We use different Markov Decision Processes (MDPs) for our tariff and wholesale market strategies. Though the MDPs are motivated by Cuevas et al. and Urieli and Stone, the novelty lies in their reward structure, solution, and application of those solutions. These are supplemented by a Neural Network based usage predictor, that also utilizes weather data. Our broker, VidyutVanika, was the runner-up in Power TAC 2018 Finals. We illustrate the efficacy of our strategies through different statistics from the competition and by conducting controlled offline experiments.

## 1.1 Contributions

The thesis makes the following contributions in the fields of smart grids, autonomous agents, game theory, and machine learning applications:

- **Double Auction Analysis**: A game theoretic analysis of single shot double auctions with a specific clearing price and payment rule, 'Average Clearing Price Rule' (ACPR), used in Power TAC

wholesale market PDAs. Specifically, we find Nash Equilibrium (NE) for One Buyer and One Seller (OBOS) and Two Buyer and One Seller (TBOS) single shot double auctions with ACPR, analytically. We also show that it is a best response to bid as close as possible to the last clearing bid in order to procure the full quantity in a generic double auction.

- **The MDPLCPBS Algorithm**: Leveraging the double auction analysis, we model the Power TAC wholesale market PDAs as MDP. Based on that, we propose an MDP based bidding strategy for PDAs, namely 'MDP and LCP based Bidding Strategy' (MDPLCPBS). MDPLCPBS has three parts - Limit Price Predictor (LPP), Quantity Predictor (QP) and Last Cleared Price Predictor (LCPP). While LPP uses MDP and dynamic programming to determine limit prices for bids, QP and LCPP are heuristics which influence the bidding quantity and dynamic programming solution of LPP respectively. The novelty lies in defining the reward functions for the MDP, solving the MDP, and applying the MDP solution to actions in the auctions. We empirically showed that MDPLCPBS follows the equilibrium derived during the double auction analysis. In addition, we benchmark our strategy against the baseline and the state-of-the art bidding strategies for the Power TAC wholesale market PDAs, and show that MDPLCPBS outperforms most of them consistently.
- **Customer Usage Predictor**: We develop a Neural Network based usage predictor for predicting customer electricity usage for future hour(s), which uses weather data and usage statistics. It is the first usage predictor to use weather data and forecast in the Power TAC setting.
- **MDP based TOU Tariff Generator**: Using an MDP formulation of the tariff market, called MDPQLM, we solve the tariff market problem using Q-learning. The novelty lies in defining the reward function for the MDP, solving the MDP, and transforming the MDP solution into a TOU tariff. We use heuristics, called Net Demand Predictor (NDP) and Tariff Designer (TaD), to convert near-optimal fixed tariffs from the MDP solution into time-of-use (TOU) tariffs.
- **VidyutVanika**: We introduce our autonomous broker, VidyutVanika, which incorporates the implementation of the learning strategies discussed above. VidyutVanika finished runner-up in the Power TAC 2018 Finals. We explain the architecture of VidyutVanika in this thesis. VidyutVanika's broker binary has been released publicly for benchmarking and research purposes.
- **Performance and Analysis**: We do an empirical analysis of VidyutVanika's performance in the Power TAC 2018 Finals. We also analyze its constituent components using offline simulations.

## 1.2 Related Work

Since 2012, several research groups have benchmarked, deployed and published brokers and associated strategies using the Power TAC platform. Outside Power TAC, most of the published bidding strategies for double auctions are designed for *Continuous Double Auctions* (CDAs). Majority of them need to be modified for *Periodic Double Auctions* (PDAs). Widespread application of PDAs have not

been seen before, but with many emerging markets, like smart grid wholesale markets, call markets etc., bidding strategies for PDAs are needed. Hence, the reason for a lot of work on PDAs to happen in the recent times, even though the literature has been limited so far.

cwiBroker [14, 22] (Power TAC 2013 & 2014 Runner-up) deployed two tariff market strategies - one for duopoly, another for oligopoly markets. For duopoly markets, cwiBroker used a heuristic inspired from Tit-For-Tat strategy in Iterated Prisoner's Dilemma, to determine tariff rates based on competing tariffs. For oligopoly markets, cwiBroker generated candidate fixed-rate tariffs and estimated their future profits, and had a fallback heuristic based strategy whenever this strategy didn't perform well. Later, cwiBroker also deployed heuristics to predict the next retail price for the next fixed price tariff to offer. On the other hand, cwiBroker used a equilibrium price and equilibrium prediction to place bids in the wholesale market. cwiBroker's wholesale market strategy was the first to introduce multiple bids for a single auction, but their multiple bids strategy is significantly different from our dummy-order LCPP strategy.

Mertacor [30] implemented two types of strategies: (i) a tariff formation strategy and (ii) a tariff update strategy. Both strategies were treated as optimization problems, where the objective function mapped to the broker's maximum profit, while simultaneously retaining an adequate portion of the customers market share. Mertacor's approach involved creating 4-6 dimensional particles which represent tariffs, and solving the optimization problem (using Particle Swarm Optimization (PSO) [10]) by finding tariffs which were predicted to give highest profit.

AstonTAC [20, 21] used Non-Homogeneous Hidden Markov Models (NHHMM) for the wholesale market to predict energy demand and clearing price, which were then fed to an MDP to determine bid prices. AstonTAC used a separate SMDP [46] to generate tariffs for the tariff market. CrocodileAgent [2, 12, 23] (3rd place Power TAC 2018) used a variant of Roth-Erev reinforcement learning algorithm to coordinate wholesale bidding across different markets by choosing among four implemented wholesale strategies. For the tariff market, CrocodileAgent used heuristics to determine hourly tariffs, and produced TOU tariffs. Maxon [49], Power TAC 2015 & 2016 Winner, used a Hill Climbing approach to generate TOU tariffs in the tariff market. Maxon also employed a tariff improvement heuristic to modify its TOU tariff as the game progressed. Furthermore, Maxon used a multiple linear regression model to predict the amount of needed energy. Past Power TAC participants, TugaTAC [42], used a fuzzy logic based mechanism to compose tariffs based on its customers portfolio. The fuzzy sets allowed it to have adaptive configurations for brokers in different scenarios.

AgentUDE [31, 32, 33], Power TAC 2014, 2017 & 2018 Winner, previously used an empirically tuned, heuristic based tariff strategy that bound customers with early withdrawl penalties and provoked competitors to reduce prices, so that customers would withdraw and pay the withdrawl penalties. In the later versions, AgentUDE used a Genetic Algorithm based strategy and aggressive pricing to design tariffs for the tariff market. In the wholesale market, AgentUDE used adaptive Q-learning to solve an MDP and determine limit prices for bidding. AgentUDE also predicted the demand of customers using a combination of SARIMA and AR models.

SPOT [5, 6] used an MDP & Q-Learning based tariff market strategy by incorporating the market share and cash position of the agent into the state space while taking actions on maintaining, incrementing or decrementing tariff rates. On the other hand, SPOT predicted bid prices for the wholesale market PDAs using REPTree, Linear Regression and neural network with weather data. In their next iteration, SPOT [7] used their REPTree based market clearing price predictor, along with a few heuristics, to design a Monte Carlo Tree Search (MCTS) based PDA bidding strategy. Their MCTS based strategy is the state-of-the-art strategy for PDAs at the time of writing, and we benchmark our MDPLCPBS algorithm's performance against it in this thesis.

COLDPower [9] (Power TAC 2016 Runner-up), inspired from Reddy and Veloso (2011), used an MDP-based strategy to generate fixed price tariffs (FPTs). This formed the base of our learning strategy, MDPQLM, for tariff markets. We modified the reward structure of the MDP, which we describe in Section 4.1. We also applied two heuristics, namely NDP (Section 4.2) and TaD (Section 4.3), on top of the FPTs, in order to generate TOU tariffs. Based on Reddy and Veloso's MDP formulation, Yang et al. (2018) presented a Recurrent Deep Multiagent Q-Learning framework with sequential clustering, to determine fixed production and consumption tariffs. They show that their Deep Q-Network (DQN) model performs better than the tabular Q-learning approach used by Reddy and Veloso.

TacTex [50, 51, 52], Power TAC 2013 Winner & 2015 Runner-up, used an MDP and dynamic programming based strategy, derived from Tesauro and Bredin's bidding strategy, to predict bid prices. Their MDP formulation was the motivation for our wholesale strategy, namely MDPLCPBS. We modified the reward function of the MDP, and while TacTex used market clearing price for their dynamic programming solution, we used our predicted last clearing price in the dynamic programming framework to arrive at optimal bid prices. This formed our Limit Price Predictor (LPP). We predicted the last clearing price using our Last Cleared Price Predictor (LCPP), which is not used by TacTex. We also spread out our bids across all the available bidding opportunities, using our Quantity Predictor (QP) strategy. Meanwhile, TacTex used a Linear Weighted Regression (LWR) based tariff market strategy which chose the best possible candidate tariff after estimating its long-term utility. Urieli and Stone also presented the design and optimization of TOU Tariffs from their LWR based FPTs, but it is significantly different from our approach.

None of the published wholesale market strategies are backed up by game theoretic analysis, where as our work on the wholesale market was to build strategies derived from Nash Equilibrium. Moreover, none of the past publications use neural networks with weather data to predict customer usage.

## 1.3 Overview

- **Chapter 2** explains the Power TAC simulation environment and the Power TAC competition.
- **Chapter 3** discusses our learning strategies for the wholesale market. This chapter aims to solve two problems for a broker in the wholesale market - (i) amount of electricity to procure/sell in the wholesale market, and (ii) bidding policy to follow while bidding in wholesale market PDAs. This

chapter does a game-theoretic analysis of single shot double auctions, and also derives the best response for a bidder with complete information in a double auction with ACPR. It also introduces the MDP based bidding strategy called MDPLCPBS and shows that it follows the game-theoretic analysis. It also benchmarks MDPLCPBS against other state-of-the-art strategies for PDAs, and shows that it beats most of them on a consistent basis.

- **Chapter 4** introduces our learning strategies in the tariff market, which use MDP and Q-Learning based approach to generate time-of-use (TOU) tariffs.
- **Chapter 5** introduces VidyutVanika, our fully autonomous broker agent, which uses our Reinforcement Learning based strategies discussed in Chapter 3 and Chapter 4. VidyutVanika was the runner-up agent in Power TAC 2018 Finals. This chapter discusses its architecture and implementation, and analyzes its performance in Power TAC 2018 Finals. It also benchmarks its submodules using controlled offline experiments to show their efficacy.
- **Chapter 6** summarizes the contributions of our work.

*Chapter 2*

# Application Domain: The Power TAC Simulator

This chapter focuses on the simulation environment used in this thesis: The Power Trading Agent Competition (Power TAC) simulation environment. The Power TAC simulation environment is open-sourced, and available on GitHub [1]. This chapter will focus only on the main components of the environment which are essential for understanding the rest of the thesis. For the detailed specifications of the various components of Power TAC, please refer to the official Power TAC game specification [19].

In Power TAC, multiple teams deploy autonomous brokers which compete against each other across three electricity markets in a simulated smart grid environment, in order to generate the highest profit.

## 2.1 Power TAC Overview

### 2.1.1 Annual Power Trading Agent Competition

The Power Trading Agent Competition (Power TAC) is an annual competition, where autonomous brokers from research groups across the globe compete against each other. The competition is divided into three phases - Trials, Qualifiers, and Finals. Each phase involves a set of games, with different configurations. A *configuration* refers to a preset number of brokers simultaneously participating in a single game. During each game, the autonomous brokers connect to the tournament server hosting the game over the internet, and communicate game actions and updates using messages.

### 2.1.2 Power TAC Simulation Environment: Overview

The Power TAC simulation environment uses state-of-the-art models to simulate a smart grid environment in a medium sized city. The simulation uses real world weather data and forecast from different locations of the world, spread across different times of the year, which affects the grid's electricity usage. Apart from modeling traditional electricity generating companies, the simulation also models renewable energy sources like wind and solar energy producers. Figure 2.1 depicts the major elements involved in the Power TAC simulation environment.

[1] `https://github.com/powertac`

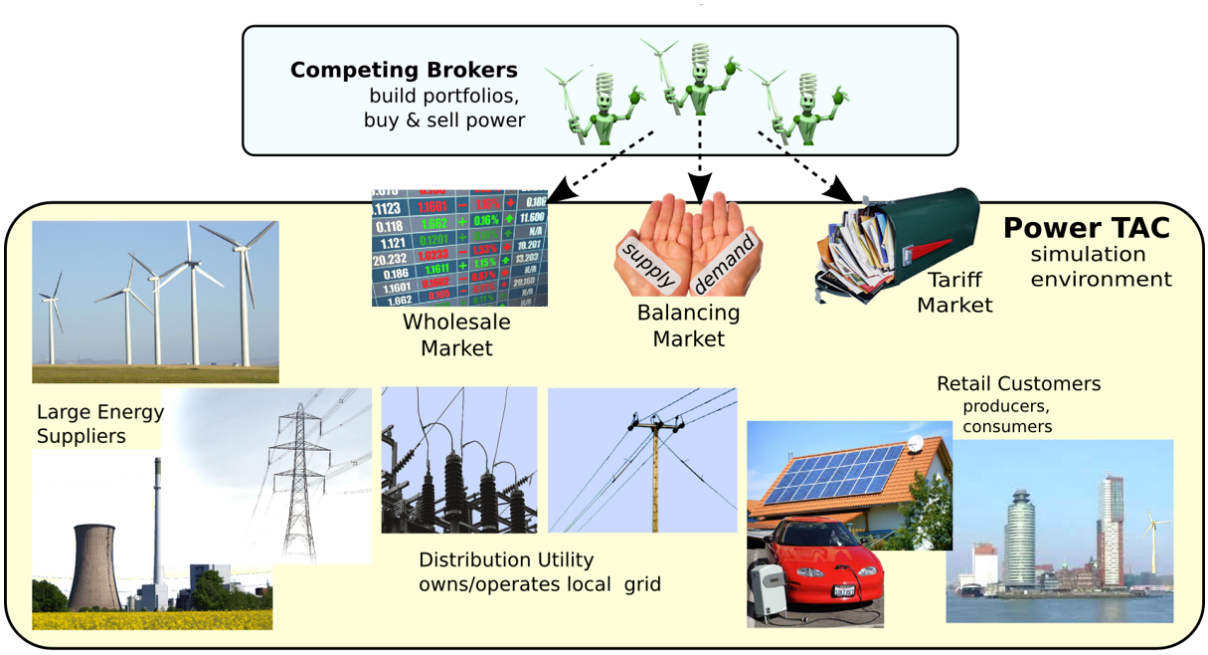


Figure 2.1: Major elements of the Power TAC simulation. **Source: The Power TAC Specification [19]**

A *game* in the Power TAC simulation environment simulates about 60 days, proceeding in 1-hour increments, also termed as *timeslots*. So there are roughly $60 \times 24 = 1440$ timeslots in each game. Each timeslot takes about 5 seconds of real time, and a full game takes about 2 hours of real time. During a game, brokers compete against each other by undertaking actions across three different markets - (1) tariff market, (2) wholesale market, and (3) balancing market. Brokers develop a subscriber base by offering bilateral trade contracts to customers in the *tariff market*, and purchase from, or sell power to this subscriber base. Meanwhile, they simultaneously attempt to fulfill their subscribers' energy requirements by trading in the *wholesale market*, in order to minimize electricity imbalance in their portfolio. Finally, any real time supply-demand imbalance within a broker's portfolio is rectified through the *balancing market*.

Power TAC simulates a population of around 57000 customers, comprising of three main power-types - *consumers*, *producers*, and *storage*. It simulates about 50000 consumers which include housing complexes, offices, hospitals and villages. A subset of these consumers accept curtailment of their electricity usage in exchange for discounted tariffs. On the other hand, around 7000 producers are simulated who use renewable sources such as solar or wind to generate electricity. Only a few comprise of the storage customers, who possess storage capacity in the form of batteries or electric vehicles connected to the smart grid. All customers have electronic appliances whose behavior and quantity are randomly selected during the initiation of each game, and not revealed to the brokers. The customers are modeled as autonomous agents who optimize their utility based on their electricity usage or production, the associated cost, and inconvenience.

In the *tariff market*, brokers attempt to draft customers into their portfolio by offering attractive tariffs that are power type specific. Tariffs could be offered with fixed or variable rates, could be tiered or based on time of use, and could include bonuses and/or fees. Since customers try to maximize their own utility, they tend to subscribe to tariffs which decrease their energy costs (or increase their profits if they are producers), and minimize their inconvenience. Inconvenience can refer to situations where customers need to shift their usage to incur lower energy costs, or continuously switch between tariffs.

The Power TAC *wholesale market* is a 'day-ahead' market, implemented using *Periodic Double Auctions* (PDAs). During these auctions, brokers interact with each other as well as with a single generating company (GenCo) and wholesale buyers, both of which are part of the simulation environment. The GenCo participating in these auctions produces roughly 11 times the net demand of the entire city i.e. all of the customers. At any given timeslot $t$, 24 independent PDAs execute in parallel for buying/selling power for the next 24 timeslots $\{t+1, \ldots, t+24\}$.

The real-time load balancing in the grid is achieved through the *balancing market*. The balancing market incentivizes brokers to balance their own portfolio, as acquiring energy via the balancing market is expensive compared to the other markets. Automatic balancing actions are carried out by the market itself, and appropriate charges are levied to the brokers for their portfolio's imbalance. Brokers can exercise *economic controls* like curtailment or up-regulation in the balancing market to reduce their imbalance charges.

Power TAC's *Distribution Utility* (DU) represents the electric utility entity that owns and operates the distribution grid. The DU connects the wholesale transmission grid to brokers and customers, while also acting as the *default broker* during the simulation. The 'default broker' offers fixed unattractive tariffs and regulates the profits of other competing brokers, as customers are free to choose these tariffs if the competing brokers offer even more unattractive tariffs. Before the start of each game, a 14-day simulation plays out with only the 'default broker', which has all the customers subscribed to it. This phase is called the *bootstrap phase*, and the data generated during this phase is sent to all the competing brokers at the start of the simulation. This bootstrap data contains the name, characteristics and consumption profile of all tariff market customers, wholesale market data pertaining to average cleared price and quantity, and weather data of the geographical location of the customer base, all at an hourly frequency.

The DU is also responsible for levying two important fees for the maintenance of the grid - (1) distribution fees, and (2) transmission capacity fees. *Distribution fees* are a fixed charge per timeslot for each customer in a broker's portfolio. There are two categories of customers: small and large. Small customers include households, electric vehicles, and offices, while large customers include industrial and warehouse models, hospitals, and larger retail producers. *Transmission capacity fees* (also termed as capacity charges), on the other hand, depict the operational cost of the grid to deliver electricity. The cost of grid's capacity to deliver the maximum demand of electricity, at a single point of time, depends on the transmission infrastructure. This bound on the capacity of the grid, and the cost associated with it, is driven by the peak demand. Thus, the DU levies capacity charges on the brokers for their customer's

contribution to the peak demand. While peak demand charges are levied retrospectively in the real world, citing a relatively short simulation time, the Power TAC simulation environment levies these charges after every week (168 timeslots). The charges are calculated as follows: 1) For each assessment timeslot $t$, compute the mean $d_t$ and standard deviation $\sigma_{d,t}$ of the net demand $d$ over all timeslots back to the start of the bootstrap phase. Compute the peak threshold defined as $z_t = \bar{d}_t + \gamma\sigma_{d,t}$ where $\gamma = 1.3$ fixed throughout the tournament. 2) Find 3 highest demand events over the last week (past 168 timeslots). For each identified peak $p > z_t$, a capacity charge $\phi_b$ is levied to each broker $b$ in proportion to their customers' contributions to the peak. The total capacity charge $\phi$ across all brokers is weighted by the amount by which the peak exceeds the threshold as $\phi = \lambda(p - z_t)$, $\lambda$ is a fixed parameter for the tournament.

The Power TAC game is competitive, dynamic, and stochastic. The state of a game is high-dimensional and rich, as there is a lot of information available to the brokers at any point during the game (see Section 2.2.4). It is also partially observable to the brokers, as the feedback from the environment can be private information for a specific broker. For example, a broker can see all the uncleared bids in the market, but doesn't know the identity of the bidders for each uncleared bid. Also, a broker can see all the published tariffs of other brokers in the market, but cannot see the corresponding subscribers of each tariff (apart from its own). The action space for brokers is also high-dimensional, as we'll see in the following section.

## 2.2 Broker Interactions with Power TAC Environment

Autonomous brokers interact with the Power TAC environment by taking actions across different markets, analyzing the feedback they got for taking those actions, and evolving their decision making process. This section describes the actions available to a broker and their corresponding feedback in the Power TAC environment. Figure 2.2 depicts the broker interaction per timeslot with the Power TAC simulation environment.

### 2.2.1 Tariff Market

During a game, in order to build a portfolio of customers, brokers can publish, revoke, and supersede tariffs in the tariff market. New and updated tariffs are only published every 6 timeslots, while customers are free to switch between tariffs at any timeslot. Tariffs can be of three main powertypes - *consumption tariff* for the consumption customers, *production tariff* for the production customers, and *storage tariff* for the battery storage tariffs. Tariffs can be more specific about the powertypes - like tariffs only for wind producers, or electric vehicles. However, tariffs cannot be published exclusively to only one customer - a tariff is applicable to customers belonging to that tariff's powertype. As long as a tariff is active (i.e. published and not revoked), customers can subscribe to it, consume or produce energy and pay or earn according to the rates of the tariff. Tariffs can be fixed priced (FPT), time-of-use (TOU), variable, tiered, or a combination of all of these. Tariffs can also include signup and early withdrawal

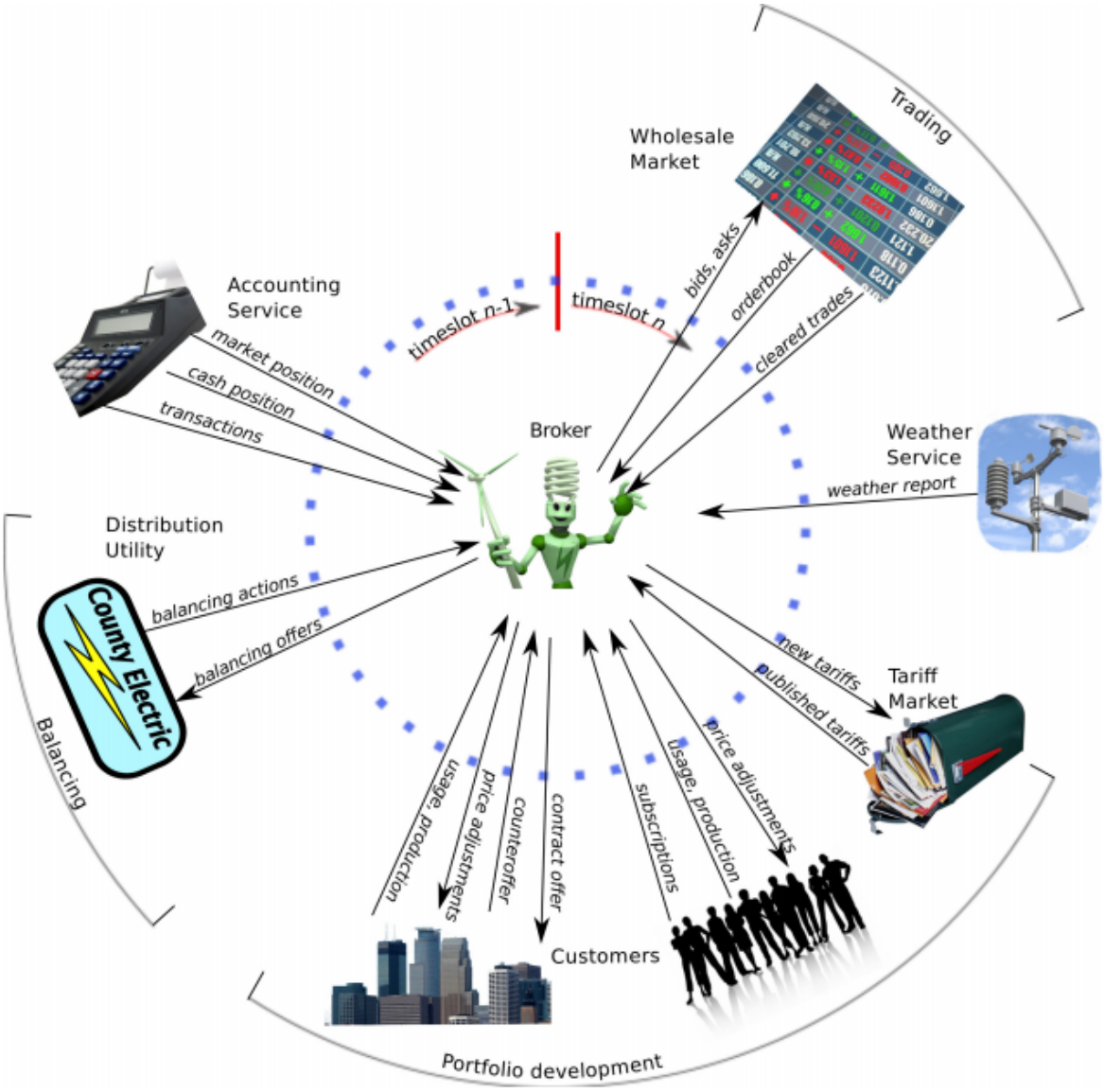


Figure 2.2: Broker Interaction with the Power TAC simulation. **Source: The Power TAC Specification [19]**. Outgoing arrows from the broker indicate possible actions by the broker, and incoming arrows indicate the feedback from the simulation environment. Counter-offers by the customers, as depicted in the figure, have not been implemented in the simulation yet. *Cash position* refers to the broker's bank balance, while *market position* refers to a broker's future energy commitments from the wholesale market i.e. the energy already contracted to be sold/bought in the future from the wholesale market.

fees/bonuses. A broker can publish/revoke any number of tariffs, but each publish or revoke action is associated by a fixed charge, which is communicated at the start of the game and varies across the tournament. This ensures that brokers do not spam the market with tariffs. Whenever a customer evaluates tariffs, it only consider five latest tariffs per broker per applicable powertype.

### 2.2.2 Wholesale Market

Brokers participate in the wholesale market by submitting bids/asks in the PDAs. The orders in the wholesale market are of the form $(quantity, price)$, where $quantity$ is the amount of energy to be bought/sold, and $price$ refers to the maximum price the broker is willing to pay per unit of energy. A bids refer to brokers wanting to buy power, while asks refer to brokers wanting to sell power. Note that in Power TAC, from the simulation perspective, bids specify a positive energy $quantity$ and a negative $price$, and asks specify a negative energy $quantity$ and a positive $price$. The sign convention is inverted from the perspective of a broker. If the $price$ is set to be NULL, then the bid/ask is considered to be a *market order* i.e. the broker is willing to pay any price for the ordered energy. Otherwise, the bid/ask is called a *limit order*.

Upon submission of bids and asks from the wholesale market players, each of the 24 PDAs execute a clearing mechanism to clear the market. Figure 2.3 showcases an example of the clearing mechanism for wholesale market PDA. In the example, there is no unique price point where the supply and demand curves intersect (which happens almost every time). As a result the mechanism dictates the average of the last executed/cleared bid (bid 8) and last executed/cleared ask (ask 6) to be the *clearing price*, which comes out to be 16. All bids higher than the last cleared bid and all asks lower than the last cleared ask are fully cleared for the same clearing price. The last cleared ask in this example is partially executed. A total of 27 MWh is cleared, and this is referred to as the *total cleared quantity*.

Post clearance, the clearing price and total cleared quantity is communicated to all the brokers. All the bids/asks which got cleared, are executed at the clearing price and corresponding brokers are notified privately. The set of uncleared bids/asks is referred to as the *orderbook*, and the orderbook is made public to all the brokers. All the bids and asks are then removed to prepare for the next auction.

### 2.2.3 Balancing Market

Brokers can indirectly place balancing orders in the balancing market by offering tariffs which specify up and down regulation rates. Brokers can also exercise economic controls on subscribed interruptible consumption customers. The strategies described in this thesis do not work with such tariffs or controls, and thus, do not take any such action.

After each timeslot, each broker portfolio's imbalance is penalized by the balancing market, and brokers are notified of their corresponding *balancing fees*. The simulator incentivizes the brokers to keep their imbalance to a minimum by penalizing the brokers heavily with high imbalance fees.

### 2.2.4 Feedback from Simulation Environment

Brokers in the Power TAC simulation get feedback of their actions via messages from the environment. Some messages are public (sent to all brokers), while some are private (for a particular broker). This section closely follows Section 3.3 of the Power TAC specification [19]. The following information is made *public* at *the start of each game*:

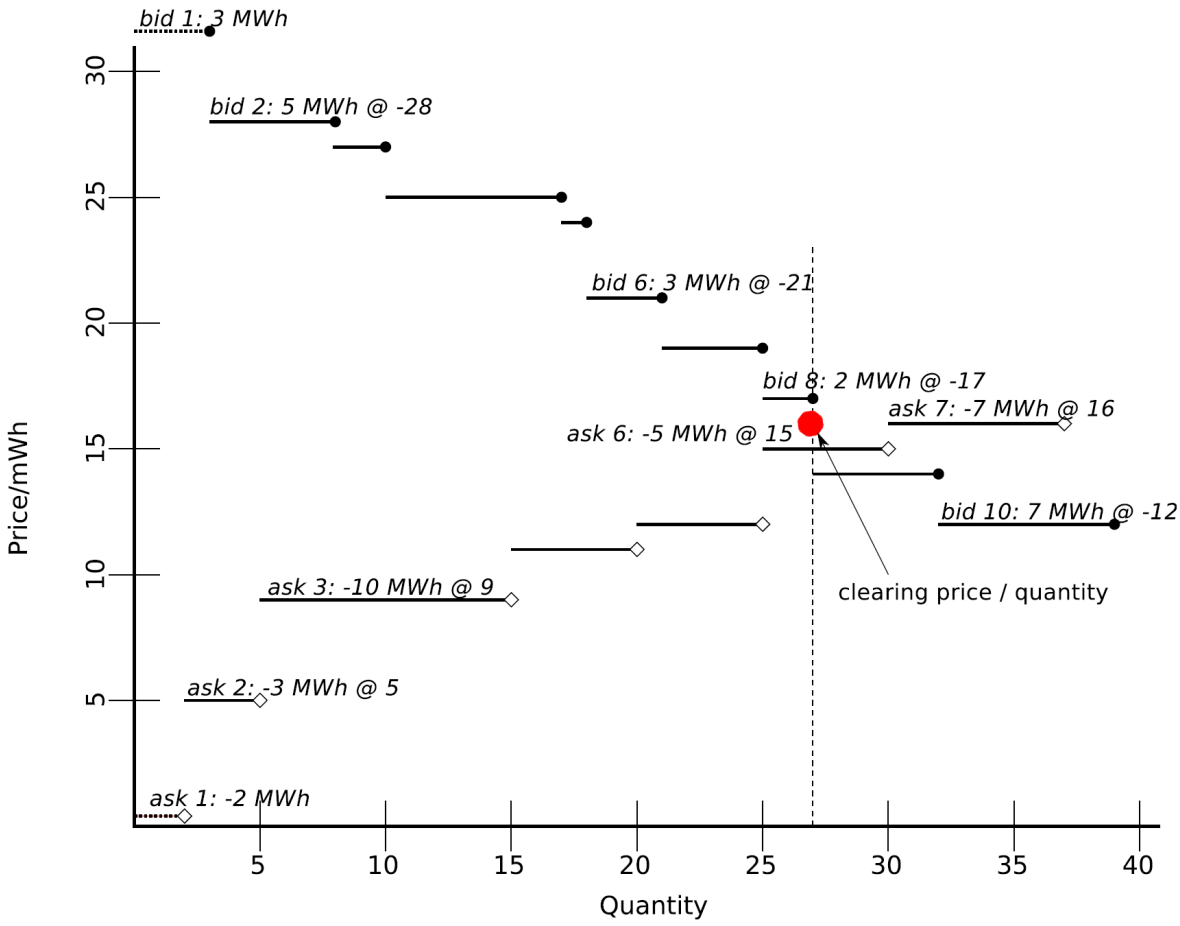


Figure 2.3: Wholesale market PDA clearing example. **Source: The Power TAC Specification [19]**

- **Game Parameters**: The parameters used to configure the current game. The parameter values for Power TAC Finals can be found in Section 8.1 of the Power TAC Specification [19].
- **Broker Identities**: The identities of all competing brokers in that particular game.
- **Customers List**: Names and properties of customers in the current game, most importantly their powertype.
- **Default Tariffs**: Tariffs offered by the default broker, to which customers are subscribed at the beginning of the game, and can re-subscribe at any point in the game. There are two such tariffs, one for producers and one for consumers.
- **Bootstrap Customer Data**: Consumption and production of each customer during bootstrap phase, under the default tariffs.
- **Bootstrap Wholesale Market Data**: Total cleared quantity and clearing price of the wholesale market during the bootstrap period. This is a result of default broker bidding in the wholesale market.

- **Bootstrap Weather Data**: Weather reports for the entire bootstrap phase.

The following information is communicated *publicly every 6 timeslots*:

- **Tariff Updates**: New tariffs, revoked tariffs and superseding tariffs submitted by all brokers.

The following information is communicated *privately every 6 timeslots*:

- **Tariff and Subscription Changes Transactions**: Tariff publication/revocation fees, and customer subscription changes (signup or withdraw) and associated charges/bonuses (early-exit penalty or singup bonus).

The following information is communicated *publicly every timeslot*:

- **Wholesale Market Clearing Data**: Clearing prices and total cleared quantities for each of the 24 PDAs in the wholesale market.
- **Wholesale Market Orderbooks**: Orderbooks (uncleared bids and asks) for each of the 24 PDAs in the wholesale market.
- **Total Energy Production and Consumption**: Total energy production and consumption for the current timeslot.
- **Weather report and Weather Forecast**: Weather conditions for the current time slot, and forecast for the next 24 timeslots.

The following information is communicated *privately every timeslot*:

- **Balancing Transactions**: Broker imbalance amount and associated charge/credit from the balancing market.
- **Portfolio Usage and Payment Transactions**: Subscribed customer usage records for the past timeslot, and associated charges/payments according to subscribed tariffs.
- **Distribution Transactions**: Broker's energy distribution quantity among it's subscribed customers and associated charges levied by DU.
- **Wholesale Market Transactions**: Cleared or partially-cleared bids and asks submitted by the broker.
- **Wholesale Market Positions**: Energy commitments made by the broker in the wholesale market which specify the energy to be delivered by/to the broker in future timeslots.
- **Cash position**: Broker's current bank balance.

The following information is communicated *publicly every 168 timeslots (1 week)*:

- **Threshold Demand and Peak Timeslots with Demand**: The threshold demand for capacity charge assessment, and the top 3 peak net demands in the market along with their respective timeslots.

The following information is communicated *privately every 168 timeslots (1 week)*:

- **Capacity Charges**: Transmission capacity fees and the associated amount by which the broker exceeded the threshold demand in the 3 peak demand timeslots (if any).

## 2.3 Summary

Even though smart grids are the future of electricity distribution, and facilitate efficient management of renewable sources of energy, there are still multiple challenges in the operationalization of smart grids and autonomous brokers operating in those smart grids, like managing highly fluctuating supply-demand scenarios, engaging stakeholders with ulterior motives, and handling automation failures of participating entities. In order to foresee such problems and examine potential solutions, Power TAC provides an open source smart grid simulator. We use the Power TAC simulator as our test-bed throughout this thesis.

This chapter describes the Power TAC simulator and the annual Power TAC tournament. In Power TAC, multiple teams deploy autonomous electricity broker agents which have to operate in the smart grid and compete in three smart electricity markets, namely tariff, wholesale and balancing market. A broker agent in Power TAC develops a subscriber base by offering attractive bilateral tariff contracts, and simultaneously attempts to fulfill its subscribers energy requirements by trading in the wholesale market. Typically, a broker agent performs three functions, (i) purchase from, or sell power to, its subscriber base in the retail (or tariff) market; (ii) purchase or sell power in the wholesale market; and (iii) rectify any supply-demand imbalance within its portfolio through the balancing market. In this chapter, we detailed the important components of the Power TAC simulator required to understand this thesis. The full specifications can be found in the Power TAC Game Specification [19]. In the following chapters, we'll describe our learning strategies for the different markets in the Power TAC simulation.

*Chapter 3*

# Learning Strategies for the Wholesale Market

This chapter focuses on learning strategies for the wholesale market. During a Power TAC game, a broker has to solve two main problems with respect to the wholesale market - (i) estimate the total quantity of energy to be purchased / sold, and (ii) the corresponding price to bid in the PDAs. Section 3.1 tackles the first problem by predicting the expected usage of the broker's portfolio using Neural Networks. To solve the latter, Section 3.2 presents a game theoretic analysis of one shot double auctions, followed by a bidding strategy for PDAs which closely follows the results from the theoretical analysis of one shot double auctions.

## 3.1 Customer Usage Predictor (CUP)

The Power TAC environment provides a weather report, and a weather forecast prediction for the next 24 time-slots. The weather report and each weather forecast has four parameters - temperature, wind direction, wind speed, and cloud cover. Since the time of the day, day of the week and the weather conditions affect the behavior of the customers, this module tries to use these 6 parameters to predict a broker's net demand.

CUP predicts the net usage of the broker's tariff portfolio for a future target time-slot $t$, by summing over the predicted usage of each customer subscribed to the broker for that target time-slot $t$. To predict the usage of each customer, it uses a Neural Network (NN) with two hidden layers of size 7 each, and 10 epochs of training over the training data. The input data consists of the weather report, time of day (0-23), and day of week (1-7), while the target variable is the actual usage of the customer. During prediction, the weather forecast is used in place of the weather report to predict the usage for the next 24 hours. A fresh model is initialized every game for each customer, and then trained on the 336 data points obtained from the bootstrap data. The model is then continuously updated via online training throughout the game, as the broker gets more data points from the usage reports for each subscribed customer.

The size of the NN is kept relatively small. This is because the response time is very small - each time-slot in the Power TAC simulation equates to just 5 seconds of real time, and the broker has to submit bids for the auctions in the next time-slot during the 5 second window. Increasing the size of

the network (layers or nodes), or training time (epochs) leads to a increase in the response time, which is not feasible. Offline training also doesn't offer convincing results as the customer behavior changes from game to game, and also depends on the subscribed tariff and time of the year - factors which we've not taken into consideration in this prediction module. The efficiency of CUP is discussed in Chapter 5.

## 3.2 Bidding in Double Auctions

This section primarily focuses on the wholesale electricity markets, and especially the Power TAC wholesale market which uses *Periodic Double Auctions* (PDAs). Auctions are mechanisms which facilitate the buying and selling of goods/items amongst a group of agents. Double auctions are more popular when the both sides of the markets actively bid. For example, in stock markets, securities are traded through double auctions. In the New York Stock Exchange, opening prices are determined using double auctions [34]. In PDAs, the market is cleared multiple times, each after a specific time interval. Any small improvement in the cost optimization, by deploying better bidding strategies can lead to significant improvements in the profits of the bidders. Motivated by this, this section takes a formal game theoretic approach for devising bidding strategies.

Typically, double auctions clearing price and payment rules differ from market to market. Equilibrium analysis of double auctions have been explored extensively with different payment and clearing price rules [57]. Specifically, for a double auction with the clearing price and payment rule as average of the last executing bid and last executing ask (ACPR), Chatterjee and Samuelson construct a symmetric equilibrium for the case of one buyer and one seller with uniformly distributed valuations [4]. Generic equilibrium analysis for the same case, and with more buyers have not been explored [57]. This section takes up a double auction with ACPR as a case study. To that end, we assume that all the involved agents (buyers and sellers) deploy scaling based strategies and identify the Nash Equilibrium (NE) of the induced game. We believe scaling based strategies are easy to interpret and implement. The equilibrium analysis of non-linear or other complex forms are analytically difficult to compute and may not be appealing to the real users of these markets. We find NEs for One Buyer and One Seller (OBOS) and Two Buyer and One Seller (TBOS) analytically (Theorem 3.2.5 and 3.2.6). Generic equilibrium analysis of double auctions with ACPR beyond these settings is challenging.

Now, if a buyer knows all the bids in a double auction (i.e. complete information setting), we argue that for such a buyer, it is a best response to bid as close as possible to the last clearing bid in order to procure the full required energy (Proposition 3.2.7). However, in reality, buyers never have access to such information. To address this lack of information, we model the bidding process in Power TAC wholesale market PDAs as a Markov Decision Process (MDP) inspired from [50], and solve it using dynamic programming. With this, we propose a strategy *MDPLCPBS* (Algorithm 1). Though our MDP formulation is similar to [50], the novelty lies in the reward, solution and application to place bids. First, we illustrate that the MDP based strategy actually achieves the equilibrium strategy identified in

Theorem 3.2.5. Then, we conduct different experiments to compare MDPLCPBS with the following strategies: ZI [11], ZIP [48], TacTex [50], and MCTS [7].

Our analysis shows that MDPLCPBS outperforms ZI, TacTex and ZIP in all the cases, and closely matches with MCTS. Note that in these experiments, the energy to be procured is same across all the brokers, and is set as some proportion of the net demand in the Power TAC simulation tariff market. MCTS is a heuristic based bidding strategy, where as MDPLCPBS is based on the game theoretic analysis of a single shot double auction.

### 3.2.1 Defintions & Background

Consider a game $\Gamma = \langle N, (S_i)_{i \in N}, (u_i)_{i \in N} \rangle$, where $N = \{1, 2, \ldots, n\}$ is the set of players, $S_i = (s_i^1, s_i^2, \ldots)$ is the (possibly infinite) strategy set of the player $i$, and $u_i : S_1 \times S_2 \times \ldots \times S_n \to \mathbb{R}$ for $i = 1, 2, \ldots n$ are utility functions.

**Definition 3.2.1.** *(*Best Response*) Given a game $\Gamma$, the best response correspondence for player $i$ is the mapping $B_i : S_{-i} \to S_i$ defined by $B_i(s_{-i}) = \{s_i \in S_i : u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i}) \forall s_i' \in S_i\}$. That is, given a profile $s_{-i}$ of strategies of the other players, $B_i(s_{-i})$ gives the set of all best response strategies of player $i$ [27].*

**Definition 3.2.2.** *(*Nash Equilibrium*) Given a game $\Gamma$, a strategy profile $s^* = (s_1^*, s_2^*, \ldots, s_n^*)$ is said to be a Nash Equilibrium of $\Gamma$ if, $u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*) \forall s_i \in S_i, \forall i = 1, 2, \ldots, n$. That is, each player's Nash Equilibrium strategy is a best response to the Nash Equilibrium strategies of the other players [27].*

**Definition 3.2.3.** *(*Markov Decision Process (MDP)*) A Markov Decision Process (MDP) [38] is a tuple given by $M = (S, A, P, r, \gamma)$ where $S$ is the set of states, $A$ is the set of actions, $P$ is the state transition probability function, where $P(s'|s, a) = P(s_{t+1} = s'|s_t = s, a_t = a)$ is the probability that action $a$ in state $s$ at time $t$ will lead to state $s'$ at time $t + 1$, $r$ is the reward function, with $r(s, a)$ denoting the reward obtained by taking action $a$ in state $s$, and $\gamma \in [0, 1]$ is the discount factor.*

*Periodic Double Auction* (PDA) is a type of auction, for buying and selling some resource, with multiple discrete clearing periods i.e. clearing after a specific time interval. Potential buyers submit their bids and potential sellers simultaneously submit their asks to an auctioneer. Then the auctioneer matches the bids and asks, and chooses some *clearing price*, denoted as $CP$, that clears the auction [58]. The *allocation rule* determines the quantity bought/sold by each buyer/seller, while the *payment rule* determines how much each buyer/seller pays/earns for buying/selling that quantity.

We formalize the *clearing price* and *payment rule* for the Power TAC PDA (Figure 2.3), as follows:

**Definition 3.2.4.** *(*Average Clearing Price Rule (ACPR)*) In a double auction, the clearing price and payment rule is* ACPR *if each cleared bid (ask) pays (earns) the CP per unit energy, where CP is average of the lowest executed bid and the highest executed ask price.*

In Power TAC, the brokers can always participate in 24 auctions to trade energy, one auction for each of the next 24 time-slots. If a broker fails to balance its demand portfolio after all the 24 auctions in the wholesale market, the balancing market automatically supplies the energy while charging the broker a $balancing\text{-}price$ for its imbalance. The $balancing\text{-}price$ is comparatively higher than the wholesale market price, and is meant to penalize the broker for having an imbalance.

### 3.2.2 Theoretical Approach and Proofs

In this section, we focus solely on the best response and NE analysis of double auctions.

#### 3.2.2.1 Nash Equilibrium analysis in single unit Double Auctions

Consider a single unit double auction, with the *clearing price* and *payment rule* given by *ACPR*. To find a generic Nash Equilibrium in this setting, we first try to simplify the double auction by restricting the number of buyers and sellers and their behavior. Upon doing so, we derive the following case-wise results.

**3.2.2.1.1 One buyer and One Seller (OBOS)** Let's assume that one buyer and one seller participate in the double auction. Let their true types be denoted by $\theta_B$ and $\theta_S$ respectively. Let us assume that the buyer follows a bidding strategy given by $b_B = \alpha_B \theta_B$, and seller follows a bidding strategy given by $b_S = \alpha_S \theta_S$, where $\alpha_B$ and $\alpha_S$ are the scale factors by which the buyer and seller scale their true types while bidding, respectively. Motivated by the literature [28] [41] [55], we choose scale based bidding strategies for this Nash Equilibrium analysis, as compared to additive bidding strategies.

Assuming an *uniform distribution* of the true types, let $h_S$ and $l_S$ denote the maximum and minimum values of $\theta_S$ respectively. Similarly, we define $h_B$ and $l_B$ as maximum and minimum values of $\theta_B$, respectively. These upper and lower limits are public information and known to both the buyer and the seller. We also assume Equation (3.1), which states that the buyer's bid (seller's ask) at any point will be less (higher) than or equal to the highest (lowest) possible seller's ask (buyer's bid).

$$\frac{\alpha_B}{\alpha_S}\theta_B \leq h_S, \qquad \frac{\alpha_S}{\alpha_B}\theta_S \geq l_B \tag{3.1}$$

Thus, the utility of the buyer if its bid gets cleared is given as:

$$\begin{aligned} u_B &= \int_{l_S}^{\frac{\alpha_B}{\alpha_S}\theta_B} \big[\theta_B - \big(\frac{\alpha_B\theta_B + \alpha_S\theta_S}{2}\big)\big] d\theta_S \\ &= \int_{l_S}^{\frac{\alpha_B}{\alpha_S}\theta_B} \big[\big(1 - \frac{\alpha_B}{2}\big)\theta_B - \frac{\alpha_S\theta_S}{2}\big] d\theta_S \\ &= \theta_B\big(1 - \frac{\alpha_B}{2}\big)\big(\frac{\alpha_B}{\alpha_S}\theta_B - l_S\big) - \frac{\alpha_S}{4}\big[\big(\frac{\alpha_B}{\alpha_S}\theta_B\big)^2 - l_S^2\big] \end{aligned} \tag{3.2}$$

Now assuming that the buyer decides to fix its $\alpha_B$ before even seeing his own type, then his utility is given by:

$$
\begin{aligned}
U_B &= \int_{l_B}^{h_B} u_B d\theta_B \\
&= \int_{l_B}^{h_B} \Big[\theta_B\big(1-\frac{\alpha_B}{2}\big)\big(\frac{\alpha_B}{\alpha_S}\theta_B - l_S\big) - \frac{\alpha_S}{4}\big[\big(\frac{\alpha_B}{\alpha_S}\theta_B\big)^2 - l_S^2\big]\Big] d\theta_B \\
&= \big(\frac{h_B^3 - l_B^3}{3}\big)\big(\frac{\alpha_B}{\alpha_S} - \frac{3\alpha_B^2}{4\alpha_S}\big) - l_S\big(1-\frac{\alpha_B}{2}\big)\big(\frac{h_B^2 - l_B^2}{2}\big) + \frac{\alpha_S}{4} l_S^2\big(h_B - l_B\big)
\end{aligned} \tag{3.3}
$$

Now, differentiating w.r.t. $\alpha_B$ and equating to 0 to find maxima

$$
\begin{aligned}
&\frac{\partial U_B}{\partial \alpha_B} = 0 \\
\Rightarrow \quad &\big(\frac{h_B^3 - l_B^3}{3}\big)\big(\frac{1}{\alpha_S} - \frac{3\alpha_B}{2\alpha_S}\big) + l_S\big(\frac{h_B^2 - l_B^2}{4}\big) = 0 \\
\Rightarrow \quad &\alpha_B = \frac{2}{3} + \frac{\alpha_S l_S}{2}\big(\frac{h_B^2 - l_B^2}{h_B^3 - l_B^3}\big)
\end{aligned} \tag{3.4}
$$

Similarly, for the seller we find the utility, when his type is known to him, to be -

$$
\begin{aligned}
u_S &= \int_{\frac{\alpha_S}{\alpha_B}\theta_S}^{h_B} \big[\big(\frac{\alpha_B\theta_B + \alpha_S\theta_S}{2}\big) - \theta_S\big] d\theta_B \\
&= \int_{\frac{\alpha_S}{\alpha_B}\theta_S}^{h_B} \big[\frac{\alpha_B}{2}\theta_B + (\frac{\alpha_S}{2} - 1)\theta_S\big] d\theta_B \\
&= \frac{\alpha_B}{4}\big(h_B^2 - (\frac{\alpha_S}{\alpha_B}\theta_S)^2\big) + \theta_S(\frac{\alpha_S}{2} - 1)\big(h_B - \frac{\alpha_S}{\alpha_B}\theta_S\big)
\end{aligned} \tag{3.5}
$$

Again, assuming that the seller decides to fix his $\alpha_S$ before even seeing his own type, then his utility is given by:

$$
\begin{aligned}
U_S &= \int_{l_S}^{h_S} u_S d\theta_S \\
&= \frac{\alpha_B h_B^2}{4}(h_S^2 - l_S^2) - \frac{\alpha_S^2}{12\alpha_B^2}(h_S^3 - l_S^3) \\
&+ h_B(\frac{\alpha_S}{2} - 1)(\frac{h_S^2 - l_S^2}{2}) - (\frac{\alpha_S}{2} - 1)\frac{\alpha_S}{\alpha_B}(\frac{h_S^3 - l_S^3}{3})
\end{aligned} \tag{3.6}
$$

Now, differentiating w.r.t $\alpha_S$ and equating to 0 to find maxima

$$
\begin{aligned}
&\frac{\partial U_S}{\partial \alpha_S} = 0 \\
\Rightarrow \quad &(\frac{h_S^3 - l_S^3}{3})\big[-\frac{\alpha_S}{2\alpha_B} + \frac{1}{\alpha_B} - \frac{\alpha_S}{\alpha_B}\big] + h_B(\frac{h_S^2 - l_S^2}{4}) = 0 \\
\Rightarrow \quad &\alpha_S = \frac{2}{3} + \frac{\alpha_B h_B}{2}(\frac{h_S^2 - l_S^2}{h_S^3 - l_S^3})
\end{aligned} \tag{3.7}
$$

Next, simplifying the expressions for $\alpha_B$ and $\alpha_S$ by letting $\frac{h_B^2 - l_B^2}{h_B^3 - l_B^3} = x$ and $\frac{h_S^2 - l_S^2}{h_S^3 - l_S^3} = y$, we get

$$
\begin{aligned}
& \alpha_S = \frac{2}{3} + \frac{\alpha_B h_B y}{2} \\
\Rightarrow \quad & \alpha_S = \frac{2}{3} + \frac{h_B y}{2}(\frac{2}{3} + \frac{\alpha_S l_S x}{2}) \\
\Rightarrow \quad & \alpha_S = \frac{4}{3}\left(\frac{2 + h_B y}{4 - l_S h_B x y}\right)
\end{aligned} \tag{3.8}
$$

$$
\begin{aligned}
& \alpha_B = \frac{2}{3} + \frac{\alpha_S l_S x}{2} \\
\Rightarrow \quad & \alpha_B = \frac{2}{3} + \frac{l_S x}{2}\frac{4}{3}\left(\frac{2 + h_B y}{4 - l_S h_B x y}\right) \\
\Rightarrow \quad & \alpha_B = \frac{4}{3}\left(\frac{2 + l_S x}{4 - l_S h_B x y}\right)
\end{aligned} \tag{3.9}
$$

Putting $l_S = l_B = 0$ and $h_S = h_B = 1$, we get $\alpha_S = 1$ and $\alpha_B = \frac{2}{3}$. Thus, the above discussion can be summarized as the following theorem.

**Theorem 3.2.5.** *For a single unit double auction with* ACPR*, with only one buyer and one seller, whose true types are drawn from a* $0 - 1$ *uniform distribution, if they deploy scaling based bidding strategies* $b_B$ *and* $b_S$ *which satisfy Equation* (3.1) *and fix their scaling factors* $\alpha_B$ *and* $\alpha_S$ *before seeing their true types, then* $\alpha_S = 1$ *and* $\alpha_B = \frac{2}{3}$ *constitute a Nash Equilibrium.*

**3.2.2.1.2 Two Buyers and One Seller (TBOS)** To find the Nash Equilibrium in this case, we proceed by assuming that both the buyers have the same scaling factor $\alpha_B$. The seller's scaling factor is denoted by $\alpha_S$. The two buyers have two different types denoted by $\theta_{B1}$ and $\theta_{B2}$, which are both assumed to have uniform distributions with the same $l_B$ and $h_B$ as lower and upper bounds, respectively. The seller's true type is denoted by $\theta_S$, and it is also assumed to have an uniform distribution with $h_S$ and $l_S$ as lower and upper bounds. Thus, the bidding strategy of buyers B1 and B2 are given by $b_{B1} = \alpha_B \theta_{B1}$ and $b_{B2} = \alpha_B \theta_{B2}$ respectively, and the seller's bidding strategy is given by $b_S = \alpha_S \theta_S$. We also assume Equation (3.10), which states that the first buyer's (seller's) bid at any point will be less than or equal to the highest possible seller's (buyer's) bid.

$$
\frac{\alpha_B}{\alpha_S}\theta_{B1} \leq h_S, \qquad \frac{\alpha_S}{\alpha_B}\theta_S \leq h_B \tag{3.10}
$$

First, we find the utility of the first buyer. We consider the following cases:

1. $b_{B1} \geq b_{B2}$ and $b_{B2} \geq b_S \Rightarrow \theta_{B1} \geq \theta_{B2}$ and $\theta_S \leq \frac{\alpha_B}{\alpha_S}\theta_{B2}$

   Let the utility in this case be denoted by $u_{b11}$.

$$
\begin{aligned}
u_{b11} &= \int_{l_B}^{\theta_{B1}} \Big[ \int_{l_S}^{\frac{\alpha_B}{\alpha_S}\theta_{B2}} \big[\theta_{B1} - \big(\frac{\alpha_B\theta_{B2} + \alpha_S\theta_S}{2}\big)\big] d\theta_S \Big] d\theta_{B2} \\
&= \int_{l_B}^{\theta_{B1}} \Big[ -\frac{3\alpha_B^2}{4\alpha_S}\theta_{B2}^2 + (\frac{\alpha_B}{\alpha_S}\theta_{B1} + \frac{\alpha_B l_S}{2})\theta_{B2} + (-l_S\theta_{B1} + \frac{l_S^2\alpha_S}{4}) \Big] d\theta_{B2} \\
&= \theta_{B1}^3(-\frac{\alpha_B^2}{4\alpha_S} + \frac{\alpha_B}{2\alpha_S}) + \theta_{B1}^2(\frac{\alpha_B l_S}{4} - l_S) + \theta_{B1}(-\frac{\alpha_B l_B^2}{2\alpha_S} + l_S l_B + \frac{l_S^2\alpha_S}{4}) \\
&+ (\frac{\alpha_B^2 l_B^3}{4\alpha_S} - \frac{\alpha_B l_S l_B^2}{4} - \frac{\alpha_S l_B l_S^2}{4})
\end{aligned} \tag{3.11}
$$

2. $b_{B2} \geq b_{B1}$ and $b_{B1} \geq b_S \Rightarrow \theta_{B1} \leq \theta_{B2}$ and $\theta_S \leq \frac{\alpha_B}{\alpha_S}\theta_{B1}$

   Let the utility in this case be denoted by $u_{b12}$.

$$
\begin{aligned}
u_{b12} &= \int_{\theta_{B1}}^{h_B} \Big[ \int_{l_S}^{\frac{\alpha_B}{\alpha_S}\theta_{B1}} \big[\theta_{B1} - \big(\frac{\alpha_B\theta_{B1} + \alpha_S\theta_S}{2}\big)\big] d\theta_S \Big] d\theta_{B2} \\
&= \int_{\theta_{B1}}^{h_B} \Big[ \theta_{B1}(1 - \frac{\alpha_B}{2})(\frac{\alpha_B}{\alpha_S}\theta_{B1} - l_S) - \frac{\alpha_S}{4}\big((\frac{\alpha_B}{\alpha_S}\theta_{B1})^2 - (l_S)^2\big) \Big] d\theta_{B2} \\
&= \theta_{B1}^3(\frac{3\alpha_B^2}{4\alpha_S} - \frac{\alpha_B}{\alpha_S}) + \theta_{B1}^2(-\frac{3\alpha_B^2}{4\alpha_S}h_B + \frac{\alpha_B}{\alpha_S}h_B + l_S - \frac{\alpha_B l_S}{2}) \\
&+ \theta_{B1}(-l_S h_B + \frac{\alpha_B l_S h_B}{2} - \frac{\alpha_S}{4}l_S^2) + \frac{\alpha_S}{4}l_S^2 h_B
\end{aligned} \tag{3.12}
$$

3. $b_{B1} \geq b_S$ and $b_{B2} \leq b_S \Rightarrow \theta_{B1} \geq \frac{\alpha_S}{\alpha_B}\theta_S$ and $\theta_{B2} \leq \frac{\alpha_S}{\alpha_B}\theta_S$

   Let the utility in this case be denoted by $u_{b13}$.

$$
\begin{aligned}
u_{b13} &= \int_{l_S}^{\frac{\alpha_B}{\alpha_S}\theta_{B1}} \Big[ \int_{l_B}^{\frac{\alpha_S}{\alpha_B}\theta_S} \big[\theta_{B1} - \big(\frac{\alpha_B\theta_{B1} + \alpha_S\theta_S}{2}\big)\big] d\theta_{B2} \Big] d\theta_S \\
&= \int_{l_S}^{\frac{\alpha_B}{\alpha_S}\theta_{B1}} \Big[ \theta_{B1}(1 - \frac{\alpha_B}{2})(\frac{\alpha_S}{\alpha_B}\theta_S - l_B) - \frac{\alpha_S}{2}(\frac{\alpha_S}{\alpha_B}\theta_S - l_B)\theta_S \Big] d\theta_S \\
&= \theta_{B1}^3(-\frac{5\alpha_B^2}{12\alpha_S} + \frac{\alpha_B}{2\alpha_S}) + \theta_{B1}^2 l_B(\frac{3\alpha_B^2}{4\alpha_S} - \frac{\alpha_B}{\alpha_S}) \\
&+ \theta_{B1}(-\frac{\alpha_S l_S^2}{2\alpha_B} + \frac{\alpha_S l_S^2}{4} + l_B l_S - \frac{\alpha_B l_B l_S}{2}) + (\frac{\alpha_S^2 l_S^3}{6} - \frac{\alpha_S l_B l_S^2}{4})
\end{aligned} \tag{3.13}
$$

Now assuming that the first buyer decides to fix his $\alpha_B$ before even seeing his own type, then we find the utility to be -

$$
\begin{aligned}
U_{B1} &= \int_{l_B}^{h_B} u_{B1} d\theta_{B1} = \int_{l_B}^{h_B} (u_{b11} + u_{b12} + u_{b13}) d\theta_{B1} \\
&= \int_{l_B}^{h_B} \Bigg[ \theta_{B1}^3 (\frac{\alpha_B^2}{12\alpha_S}) + \theta_{B1}^2 (-\frac{\alpha_B l_S}{4} + \frac{\alpha_B}{\alpha_S}(h_B - l_B) - \frac{3\alpha_B^2}{4\alpha_S}(h_B - l_B)) \\
&+ \theta_{B1} \Bigg( 2 l_B l_S - l_S h_B - \frac{\alpha_B l_B^2}{2\alpha_S} + \frac{\alpha_B l_S h_B}{2} - \frac{\alpha_S l_S^2}{2\alpha_B} + \frac{\alpha_S l_S^2}{4} - \frac{\alpha_B l_B l_S}{2} \Bigg) \\
&+ (\frac{\alpha_B^2 l_B^3}{4\alpha_S} - \frac{\alpha_B l_S l_B^2}{4} - \frac{\alpha_S l_B l_S^2}{2} + \frac{\alpha_S^2 l_S^3}{6} + \frac{\alpha_S}{4} l_S^2 h_B) \Bigg] d\theta_{B1}
\end{aligned} \tag{3.14}
$$

Now, differentiating w.r.t $\alpha_B$ and equating to 0 to find maxima

$$
\begin{aligned}
& \frac{\partial U_{B1}}{\partial \alpha_B} = 0 \\
\Rightarrow \quad & \Bigg[ (h_B - l_B)(\frac{\alpha_B l_B^3}{2\alpha_S} - \frac{l_S l_B^2}{4}) + (\frac{h_B^2 - l_B^2}{2}) \Bigg( - \frac{l_B^2}{2\alpha_S} + \frac{l_S h_B}{2} + \frac{\alpha_S l_S^2}{2\alpha_B^2} - \frac{l_B l_S}{2} \Bigg) \\
& + (\frac{h_B^3 - l_B^3}{3}) \Bigg( - \frac{l_S}{4} + \frac{h_B - l_B}{\alpha_S} - \frac{3\alpha_B}{2\alpha_S}(h_B - l_B) \Bigg) + (\frac{h_B^4 - l_B^4}{4})(\frac{\alpha_B}{6\alpha_S}) \Bigg] = 0 \\
\Rightarrow \quad & \Bigg[ \frac{\alpha_B}{24\alpha_S}(-11h_B^3 + 25l_B^3 + l_B^2 h_B + l_B h_B^2) + \frac{\alpha_S}{4\alpha_B^2}(h_B l_S^2 + l_B l_S^2) \\
& + \frac{l_S(2h_B^2 - 7l_B^2 - h_B l_B)}{12} + \frac{4h_B^3 - 7l_B^3 - 3h_B l_B^2}{12\alpha_S} \Bigg] = 0
\end{aligned} \tag{3.15}
$$

Similarly, for the seller we find the utility. We again have 4 cases:

1. $b_{B1} \geq b_{B2}$ and $b_{B2} \geq b_S \Rightarrow \theta_{B1} \geq \theta_{B2}$ and $\theta_{B2} \geq \frac{\alpha_S}{\alpha_B}\theta_S$

   Let the utility in this case be denoted by $u_{s1}$.

$$
\begin{aligned}
u_{s1} &= \int_{\frac{\alpha_S}{\alpha_B}\theta_S}^{h_B} \Bigg[ \int_{\theta_{B2}}^{h_B} \big[ (\frac{\alpha_B \theta_{B2} + \alpha_S \theta_S}{2}) - \theta_S \big] d\theta_{B1} \Bigg] d\theta_{B2} \\
&= \int_{\frac{\alpha_S}{\alpha_B}\theta_S}^{h_B} \Bigg[ \theta_S (\frac{\alpha_S}{2} - 1)(h_B - \theta_{B2}) + \frac{\alpha_B}{2}\theta_{B2}(h_B - \theta_{B2}) \Bigg] d\theta_{B2} \\
&= \theta_S^3 (\frac{5\alpha_S^3}{12\alpha_B^2} - \frac{\alpha_S^2}{2\alpha_B^2}) + \theta_S^2 (-\frac{3\alpha_S^2 h_B}{4\alpha_B} + \frac{\alpha_S h_B}{\alpha_B}) + \theta_S (\frac{\alpha_S}{4} - \frac{1}{2}) h_B^2 + (\frac{\alpha_B h_B^3}{12})
\end{aligned} \tag{3.16}
$$

2. $b_{B2} \geq b_{B1}$ and $b_{B1} \geq b_S \Rightarrow \theta_{B2} \geq \theta_{B1}$ and $\theta_{B1} \geq \frac{\alpha_S}{\alpha_B}\theta_S$

   Let the utility in this case be denoted by $u_{s2}$. Since the two buyers are symmetric, the utility in this case comes to be same as in case 1.

$$
u_{s2} = \theta_S^3 (\frac{5\alpha_S^3}{12\alpha_B^2} - \frac{\alpha_S^2}{2\alpha_B^2}) + \theta_S^2 (-\frac{3\alpha_S^2 h_B}{4\alpha_B} + \frac{\alpha_S h_B}{\alpha_B}) + \theta_S (\frac{\alpha_S}{4} - \frac{1}{2}) h_B^2 + (\frac{\alpha_B h_B^3}{12}) \tag{3.17}
$$

3. $b_{B1} \geq b_S$ and $b_{B2} \leq b_S \Rightarrow \theta_{B1} \geq \frac{\alpha_S}{\alpha_B}\theta_S$ and $\theta_{B2} \leq \frac{\alpha_S}{\alpha_B}\theta_S$

Let the utility in this case be denoted by $u_{s3}$.

$$\begin{aligned} u_{s3} &= \int_{l_B}^{\frac{\alpha_S}{\alpha_B}\theta_S} \left[ \int_{\frac{\alpha_S}{\alpha_B}\theta_S}^{h_B} \left[ \left( \frac{\alpha_B\theta_{B1} + \alpha_S\theta_S}{2} \right) - \theta_S \right] d\theta_{B1} \right] d\theta_{B2} \\ &= \int_{l_B}^{\frac{\alpha_S}{\alpha_B}\theta_S} \left[ \theta_S(\frac{\alpha_S}{2} - 1)(h_B - \theta_S) + \frac{\alpha_B}{4}(h_B^2 - \theta_S^2) \right] d\theta_{B2} \\ &= \theta_S^3(\frac{\alpha_S^2}{\alpha_B^2} - \frac{3\alpha_S^3}{4\alpha_B^2}) + \theta_S^2(\frac{\alpha_S^2 h_B}{2\alpha_B} + \frac{3\alpha_S^2 l_B}{4\alpha_B} - \frac{\alpha_S h_B}{\alpha_B} - \frac{\alpha_S l_B}{\alpha_B}) \\ &+ \theta_S(\frac{\alpha_S h_B^2}{4} - h_B l_B(\frac{\alpha_S}{2} - 1)) - \frac{\alpha_B h_B^2 l_B}{4} \end{aligned} \tag{3.18}$$

4. $b_{B2} \geq b_S$ and $b_{B1} \leq b_S \Rightarrow \theta_{B2} \geq \frac{\alpha_S}{\alpha_B}\theta_S$ and $\theta_{B1} \leq \frac{\alpha_S}{\alpha_B}\theta_S$

Let the utility in this case be denoted by $u_{s4}$. Since the two buyers are symmetric, the utility in this case comes to be same as in case 3.

$$\begin{aligned} u_{s4} &= \theta_S^3(\frac{\alpha_S^2}{\alpha_B^2} - \frac{3\alpha_S^3}{4\alpha_B^2}) + \theta_S^2(\frac{\alpha_S^2 h_B}{2\alpha_B} + \frac{3\alpha_S^2 l_B}{4\alpha_B} - \frac{\alpha_S h_B}{\alpha_B} - \frac{\alpha_S l_B}{\alpha_B}) \\ &+ \theta_S(\frac{\alpha_S h_B^2}{4} - h_B l_B(\frac{\alpha_S}{2} - 1)) - \frac{\alpha_B h_B^2 l_B}{4} \end{aligned} \tag{3.19}$$

Now assuming that the seller decides to fix his $\alpha_S$ before even seeing his own type, then we find the utility to be -

$$\begin{aligned} U_S &= \int_{l_S}^{h_S} u_S d\theta_S \\ &= \int_{l_S}^{h_S} (u_{s1} + u_{s2} + u_{s3} + u_{s4}) d\theta_S = 2\int_{l_S}^{h_S} (u_{s1} + u_{s3}) d\theta_S \\ &= 2\int_{l_S}^{h_S} \left[ \theta_S^3(\frac{\alpha_S^2}{2\alpha_B^2} - \frac{\alpha_S^3}{3\alpha_B^2}) + \theta_S^2(\frac{3\alpha_S^2 l_B}{4\alpha_B} - \frac{\alpha_S l_B}{\alpha_B} - \frac{\alpha_S^2 h_B}{4\alpha_B}) \right. \\ &\left. + \theta_S(\frac{\alpha_S h_B^2}{2} - h_B l_B(\frac{\alpha_S}{2} - 1) - \frac{h_B^2}{2}) + \frac{\alpha_B h_B^3}{12} - \frac{\alpha_B h_B^2 l_B}{4} \right] d\theta_S \end{aligned} \tag{3.20}$$

Now, differentiating w.r.t $\alpha_S$ and equating to 0 to find maxima

$$
\begin{aligned}
& \frac{\partial U_S}{\partial \alpha_S} = 0 \\
\Rightarrow \quad & \Big[ (\frac{h_S^4 - l_S^4}{4})(\frac{\alpha_S}{\alpha_B^2} - \frac{\alpha_S^2}{\alpha_B^2}) + (\frac{h_S^3 - l_S^3}{3})(\frac{3\alpha_S l_B}{2\alpha_B} - \frac{l_B}{\alpha_B} - \frac{\alpha_S h_B}{2\alpha_B}) \\
& + (\frac{h_S^2 - l_S^2}{2})(-\frac{h_B l_B}{2} + \frac{h_B^2}{2}) \Big] = 0 \\
\Rightarrow \quad & -\frac{\alpha_S^2}{4\alpha_B^2}(h_S^4 - l_S^4) + \frac{\alpha_S}{\alpha_B}(\frac{h_S^4 - l_S^4}{4\alpha_B} + \frac{(h_S^3 - l_S^3)(3l_B - h_B)}{6}) \\
& - \frac{l_B}{\alpha_B}(\frac{h_S^3 - l_S^3}{3}) + (\frac{h_S^2 - l_S^2}{2})(-\frac{h_B l_B}{2} + \frac{h_B^2}{2}) = 0
\end{aligned}
\tag{3.21}
$$

From Equation (3.15), we have a bi-variate cubic equation in $\alpha_B$ and $\alpha_S$, and from Equation (3.21), we have a bi-variate quadratic equation in $\alpha_B$ and $\alpha_S$.

Assuming $\alpha_S \neq 0$ and $\alpha_B \neq 0$ (non-zero bids), and putting $l_S = l_B = 0$ and $h_S = h_B = 1$ in Equation (3.15), we get

$$
\begin{aligned}
& \frac{-11\alpha_B}{24\alpha_S} + \frac{4}{12\alpha_S} = 0 \\
\Rightarrow \quad & \alpha_B = \frac{8}{11}
\end{aligned}
\tag{3.22}
$$

Now, putting $\alpha_B = \frac{8}{11}$ (from Equation (3.22)), $l_S = l_B = 0$ and $h_S = h_B = 1$ in Equation (3.21), we get

$$
\begin{aligned}
& -\alpha_S^2 + \frac{17\alpha_S}{33} + \frac{64}{121} = 0 \\
\Rightarrow \quad & \alpha_S = \frac{17 \pm \sqrt{2593}}{66}
\end{aligned}
\tag{3.23}
$$

Since $\alpha_S = \frac{17-\sqrt{2593}}{66} < 0$ (negative scaling factor), we ignore this solution.

Thus, Putting $l_S = l_B = 0$ and $h_S = h_B = 1$ in Equation (3.15) and Equation (3.21), we get $\alpha_S = \frac{17+\sqrt{2593}}{66} \approx 1.02911$ and $\alpha_B = \frac{8}{11} \approx 0.727273$. The above discussion can be summarized as the following theorem.

**Theorem 3.2.6.** *For a single unit double auction with* ACPR *with two buyers and one seller, whose true types are drawn from a* $0-1$ *uniform distribution, if they deploy scaling based strategies* $b_{B1}$*,* $b_{B2}$ *and* $b_S$*, with buyers having the same scaling factor* $\alpha_B$*, which satisfy Equation* (3.10) *and fix their scaling factors* $\alpha_B$ *and* $\alpha_S$ *before seeing their true types, then* $\alpha_S = \frac{17+\sqrt{2593}}{66}$ *and* $\alpha_B = \frac{8}{11}$ *constitute a Nash Equilibrium.*

As seen, with the increase in just one buyer, the complexity of the solution increases. It becomes increasingly difficult to extend and generalize the above results for a realistic market setting. Thus, moving forward, taking the Power TAC wholesale market as testbed, we present a bidding strategy in Section 3.2.3 and experimentally show that it follows the theoretical results obtained in this section.

#### 3.2.2.2 Best Response analysis in multi-unit Double Auctions with complete information

In practice, there are key differences between double auctions implemented in markets, and the theoretical results arrived above, stated as follows:

1. Quantity may be involved in the trading market auctions, which is not considered above.
2. The seller needs to use the same bidding strategy for one to achieve the above result, which may not the case.

So, considering a multi-unit double auction, we derive the best response if all the other bids are known to the bidder (i.e. complete information).

Given a double auction for trading energy contracts, with *ACPR*, let $b_i(pb_i, qb_i)$ denote the $i^{th}$ bid placed in the auction, for $qb_i$ amount of energy, at $pb_i$ price. Similarly, let $a_i(pa_i, qa_i)$ denote the $i^{th}$ ask placed in the auction, where $pa_i$ and $qa_i$ denote the asking price and quantity respectively. For simplicity, let us assume the ordering of bids to be in descending order of price, and ordering of asks to be in ascending order of price. Therefore, $pb_i > pb_{i+1}$, and $pa_i < pa_{i+1}$. The last clearing bid (LCB) is denoted by $b_{c_1}(pb_{c_1}, qb_{c_1})$, while the last clearing ask (LCA) is denoted by $a_{c_2}(pa_{c_2}, qa_{c_2})$. Thus, the clearing price is given as $CP = (pb_{c_1} + pb_{c_2})/2$. Let $Q_a$ denote the energy not cleared of the LCA if LCA is executed partially, and let $Q_b$ denote the energy not cleared of the LCB if LCB is executed partially. We assume that the bids are not market orders.

**Claim 1.** *Upon clearance of an auction, either $Q_a$ or $Q_b$, or both have to be zero.*

*Proof.* If the last bid partially clears, $Q_a = 0$ and $Q_b \neq 0$, and if the last ask partially clears, $Q_a \neq 0$ and $Q_b = 0$. If both clear fully, $Q_a = 0$ and $Q_b = 0$. The last bid and last ask both can't clear partially, as, if they did, then more quantity can be cleared with last bid's price higher than the last ask's price. □

**Proposition 3.2.7.** *When a buyer (seller) has complete information about the auction, and it desires to procure (sell) entire energy it bids (asks) for, it's a best response to bid as close as possible to the last clearing bid (ask).*

*Proof.* WLOG, let us consider the case of bids in the auction. Our claim essentially solves the optimization problem of minimizing the clearing price while procuring the full amount of energy. Assume a buyer $m$ wants to place a bid $b_m(pb_m, qb_m)$ in such an auction. We define $Q_a$ and $Q_b$ denote the energy not cleared of last bid and last ask respectively, when the buyer $m$ doesn't participate. Now if the buyer does participate in the auction, there are the following possibilities (depicted in Figure 3.1):

**Case 1:** $pb_m < pb_{c_1}$ i.e. bid price is lower than price of the would-be cleared bid when the buyer doesn't participate

1. $pb_{c_1} > pa_{c_2} > pb_m$ and $Q_a \geq 0$, $Q_b \geq 0$ i.e. if the bidding price of $m$ is lower than the ask price of the last cleared ask (when buyer doesn't participate). Under this condition, the bid doesn't clear, and this is clearly not optimal, as the buyer doesn't get it's required energy.

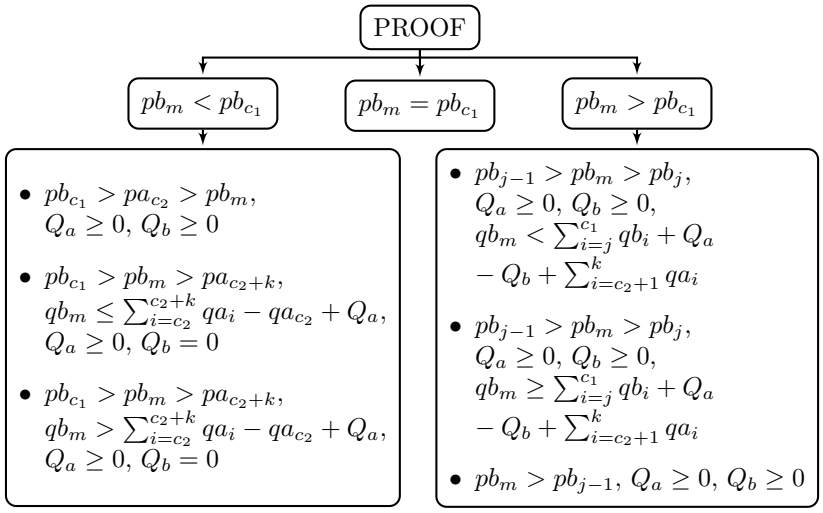


Figure 3.1: Proof Cases

2. $pb_{c_1} > pb_m > pa_{c_2+k}$, $Q_a \geq 0$ and $Q_b = 0$ and $k$ is the smallest index with $k \geq 0$ such that $qb_m \leq \sum_{i=c_2}^{c_2+k} qa_i - qa_{c_2} + Q_a$, i.e. if the last cleared bid is fully executed when buyer doesn't participate, and buyer $m$'s bid price is higher than the next closest ask. In this case, clearly, $b_m$ becomes the last clearing bid, and it gets cleared fully. Thus, buyer $m$ gets energy at the lowest price possible by having the last cleared bid.

3. $pb_{c_1} > pb_m > pa_{c_2+k}$, where $Q_a \geq 0$ and $Q_b = 0$ and $k$ is the largest index with $k \geq 0$ such that $qb_m > \sum_{i=c_2}^{c_2+k} qa_i - qa_{c_2} + Q_a$, i.e. if the last cleared bid is fully executed when buyer doesn't participate, and buyer $m$'s bid price is higher than the next ask. In this case, clearly, $b_m$ becomes the last clearing bid, and it gets cleared partially. Although the buyer $m$ gets some energy at the lowest price possible by having the last cleared bid, it would've been better off bidding higher than the previous bid $b_c$, in order to clear it's entire bid energy $qb_m$.

**Case 2:** $pb_m = pb_{c_1}$ i.e. bidding price same as the last cleared bid's price (when buyer doesn't participate). This is a probability zero event, and extremely unlikely to occur. Since it's a tie, it'll either be treated as Case 1 or Case 3, depending on the tiebreaker rule set by the auction.

**Case 3:** $pb_m > pb_{c_1}$ i.e. bidding price is just higher than the last cleared bid's price (when buyer doesn't participate)

1. $pb_{j-1} > pb_m > pb_j$ and $Q_a \geq 0$, $Q_b \geq 0$, where $j$ is the largest index with $j \leq c_1$ such that $qb_m \leq \sum_{i=j}^{c_1} qb_i + Q_a - Q_b + \sum_{i=c_2+1}^{k} qa_i$ and $k$ is the largest index such that $pa_k \leq pb_j$. In this case, $b_m$ clears fully and becomes the second last bid to clear, and $b_j$ clears partially, or $b_m$ clears fully and becomes the last cleared bid. If the buyer decides to bid below $pb_j$, it'll clear partially, which is not desirable, and thus supports our claim.

2. $pb_{j-1} > pb_m > pb_j$ and $Q_a \geq 0$, $Q_b \geq 0$, where $j$ is the smallest index such that $qb_m > \sum_{i=j}^{c_1} qb_i + Q_a - Q_b + \sum_{i=c_2+1}^{k} qa_i$ and $k$ is the largest index such that $pa_k \leq pb_j$. In this case, $b_m$ becomes the last clearing bid, and executes partially. The buyer is better off bidding higher than $b_{j-1}$ as it would've cleared fully.

3. $pb_m > pb_{j-1}$ and $Q_a \geq 0$, $Q_b \geq 0$, where $j$ is the largest index such that $qb_m \leq \sum_{i=j}^{c_1} qb_i + Q_a - Q_b + \sum_{i=c_2+1}^{k} qa_i$ and $k$ is the largest index such that $pa_k \leq pb_j$. In this case, $b_m$ clears fully. But, the buyer $m$ would get the full amount of energy even if it bid between $pb_{j-1}$ and $pb_j$, and would've then become close to the clearing bid.

□

Given the above proposition in a complete information setting, we further propose a MDP-based bidding strategy, which uses the past auction trends and statistics, to achieve the best response with incomplete information in the Power TAC wholesale market.

### 3.2.3 MDPLCPBS: Power TAC Wholesale Market Bidding Strategy

We introduce the MDP and LCP based Bidding Strategy (MDPLCPBS) for the Power TAC wholesale market. The Power TAC wholesale market accepts bids of the form $(quantity, limit\text{-}price)$. With respect to a broker, let the energy amount being sold be positive, while the energy amount being bought be negative. Meanwhile, let negative price indicate a broker is earning revenue, while positive price indicate it is paying or losing revenue. Thus, from the viewpoint of a broker, a buy order is seen to have a negative $quantity$ and a positive $limit\text{-}price$, while a sell order (termed as an *ask*), is seen to have a positive $quantity$ and a negative $limit\text{-}price$.

At time-slot $t$, assuming a broker has a predicted demand profile $D_t = \{d_{t+1}, d_{t+2}, \ldots, d_{t+24}\}$, where $d_i$ is the predicted net demand at time-slot $i$. Also, let $P_t = \{p_{t+1}, p_{t+2}, \ldots, p_{t+24}\}$ denote the amount of energy already procured by past energy contracts, where $p_i$ denotes the market position for time-slot $i$. Thus, the remaining energy to be procured is given by $E_t = \{e_{t+1}, e_{t+2}, \ldots, e_{t+24}\}$, where $e_i = d_i - p_i$ is the net energy left to be procured for time-slot $i$. The bidding strategy, MDPLCPBS, to procure the aforementioned energy requirements, comprises of three major submodules - (i) Limit Price Predictor, (ii) Quantity Predictor, and (iii) Last Cleared Price Predictor.

#### 3.2.3.1 Limit Price Predictor (LPP)

At any given time-slot $t$, the predictor computes 24 $limit\text{-}prices$ for 24 simultaneous PDAs in the Power TAC wholesale market. Motivated by [47] and [50], the Limit Price Predictor uses the following MDP to place optimal $limit\text{-}prices$ for bids:

1. **States**: $s \in S = \{0, 1, \ldots, 24, success\}$, $s_0 := 24$
2. **Actions**: $limit\text{-}price \in \mathbb{R}$

3. **Transition**: The same state transition from [50] is used. A state $s \in \{1, \ldots, 24\}$ transitions to one of two states. If a bid is partially or fully cleared, it transitions to the terminal state $success$. Otherwise, a state $s$ transitions to state $s - 1$. The clearing (i.e. transition) probability $p_{cleared}(s, limit\text{-}price)$ is initially unknown and is determined by Equation (3.25).
4. **Reward**: At any state $s \in \{1, \ldots, 24\}$, the reward is 0. At terminal state $s = 0$, the reward is the negative of the *balancing price* per unit energy. At terminal state $s = success$, the reward is the negative of the *limit-price* of the cleared bid. Since we take the price to be positive for bids and negative for asks, maximizing reward results in minimizing costs.
5. **Terminal States**: $\{0, success\}$

We solve the above MDP using a sequential bidding strategy, that computes the optimal bid $limit\text{-}price$ that minimizes the expected procurement cost per unit energy. It uses the $balancing\text{-}price$ as the expected price at state $s = 0$, and recursively minimizes the expected cost by using the probability of clearance, $p_{cleared}(s, limit\text{-}price)$. This solution is summarized as a value function, stated as follows:

$$V(s) = \begin{cases} balancing\text{-}price, & \text{if } s = 0 \\ \min\limits_{limit\text{-}price} \{p_{cleared} \times limit\text{-}price & \\ +(1 - p_{cleared}) \times V(s-1)\}, & \text{if } s \in [1, 24] \end{cases} \tag{3.24}$$

Given that the $balancing\text{-}price$ and the $p_{cleared}$ values are different for bids and asks, we maintain two separate instances of the MDP, and solve them independently.

The value function in Equation (3.24) is solved recursively using dynamic programming. However, before doing so, the $balancing\text{-}price$ and the transition function $p_{cleared}(s, limit\text{-}price)$ need to be estimated, as they are both initially unknown. The $balancing\text{-}price$ is estimated by averaging the balancing-prices across past time-slots. On the other hand, the clearing probability, $p_{cleared}(s, limit\text{-}price)$, is computed using past auction statistics as:

$$p_{cleared} = \frac{\sum_{ac \in auction[s], ac.\text{LCP} < limit\text{-}price} ac.\text{cleared-amount}}{\sum_{ac \in auction[s]} ac.\text{cleared-amount}} \tag{3.25}$$

where $auction[s]$ is the set of all past auctions in the state $s$, and LCP is the *Last Clearing Price*, which is estimated by the *Last Cleared Price Predictor* in section 3.2.3.3. The auction statistics for each state $s$ are re-used in the future for estimating $p_{cleared}$, as we iterate over the same sequence of states S during the bidding process.

#### 3.2.3.2 Quantity Predictor (QP)

The *Quantity Predictor* is primarily responsible for distributing the demand for a target time-slot across all the 24 auctions, in order to further reduce overall energy cost. The idea is to buy more and sell less at cheaper prices, and vice-versa. It essentially breaks down the demand for a target time-slot $t + 24$, across auctions in time-slots $\{t, t + 1, \ldots, t + 23\}$.

For each auction state $s \in \{1, \ldots, 24\}$ at time-slot $t$, it takes the corresponding energy requirement $e_{t+s}$ and uses the 24 *limit-prices* from the *Limit Price Predictor* to distribute the required energy. The energy quantity to bid/ask, for each state $s$ at time-slot $t$, is given by:

$$q(s) = \begin{cases} \frac{e_{t+s}}{\sum_{j=s}^{24} \frac{limit\text{-}price[j]}{limit\text{-}price[s]}}, & \text{if } e_{t+s} > 0 \\ \frac{e_{t+s}}{\sum_{j=s}^{24} \frac{limit\text{-}price[s]}{limit\text{-}price[j]}}, & \text{if } e_{t+s} < 0 \\ 0, & \text{if } e_{t+s} = 0 \end{cases} \tag{3.26}$$

where $s \in \{1, \ldots, 24\}$, $limit\text{-}price[s]$ is the limit-price for state $s$ determined by the *Limit Price Predictor*. The first case in Equation (3.26) refers to the situation where energy needs to sold, so the bid quantity is directly proportional to the predicted limit-price of that auction - essentially selling more energy at higher price. On the other hand, the second case occurs when the energy needs to be procured. So, the bid quantity is set to be inversely proportional to the predicted limit-price i.e. buying more energy at cheaper price. Thus, the final bid is of the form $(q(s), limit\text{-}price[s])$.

#### 3.2.3.3 Last Cleared Price Predictor (LCPP)

First, one has to note that, in any auction, the LCP is greater than or equal to CP. Mostly, $LCP > CP$, as $P(LCP = CP) = 0$, i.e. LCP equal to CP is a probability zero event. In Power TAC, the LCP is not known to any broker. In essence, one can place better bids if the LCP for each auction is known, as they can bid higher than a predicted LCP to become the last bid, and achieve best response according to Proposition 3.2.7. The Last Cleared Price Predictor essentially tries to determine the LCP for bids and asks for all executed auctions. It does so by probing the auctions with a set of dummy orders, which have the minimum tradeable energy as quantity (0.01 MwH), and $limit\text{-}prices$ equally spaced in the range $[\beta \times limit\text{-}price, balancing\text{-}price]$. After execution, the LCP for bids for an auction in state s is determined by:

$$LCP(s) = \min(dummy\text{-}bids_{cleared}, limit\text{-}price[s]_{cleared}) \tag{3.27}$$

where $dummy\text{-}bids_{cleared}$ is the set of bid prices of all dummy bids which got cleared in the state $s$, and $limit\text{-}price[s]_{cleared}$ is the limit-price for the cleared final bid made in state $s$ (taken to be infinity if final bid didn't clear or doesn't exist). Similarly, the LCP for asks is given as:

$$LCP(s) = \max(dummy\text{-}asks_{cleared}, limit\text{-}price[s]_{cleared}) \tag{3.28}$$

where $dummy\text{-}asks_{cleared}$ is the set of ask prices of all dummy asks which got cleared in the state $s$, and $limit\text{-}price[s]_{cleared}$ is the limit-price for the cleared final ask made in state $s$ (taken to be infinity if final ask didn't clear or doesn't exist). These LCP values are then used to update the clearing probability $p_{cleared}$ in Equation (3.25).

Algorithm 1 summarizes MDPLCPBS, which is executed every time-slot. It takes the energy requirement for the 24 auctions as input. First it collects the market statistics, which includes the LCP

**Algorithm 1** MDPLCPBS

```
1:  procedure MDPLCPBS(energyReq[1..24])
2:      marketData[0..24] ← getMarketStatistics()
3:      if EnoughDataPoints(marketData) then
4:          bidPrices[1..24] ← SolveMDP(marketData)
5:          bidQty[1..24] ← DistributeQty(energyReq, bidPrices)
6:      else
7:          bidPrices[1..24] ← SampleBiddingPolicy()
8:          bidQty[1..24] ← energyReq[1..24]
9:      end if
10:     sendBids(bidPrices, bidQty)
11:     sendDummyBids(bidPrices, marketData)
12: end procedure
```

estimate and clearing amount from previous time-slots, and the balancing price (line 2). If the number of data points is suitable enough, it proceeds to solve the MDP and generates a set of prices to bid (line 4). Using these set of prices, and the energy requirements, it generates a set of quantities to bid (line 5). If data points are not enough, the bidding policy given in the Power TAC *sample-broker* is used to determine the bid prices (line 7), and the bid quantities are set as the full energy requirements (line 8). Using the determined bid prices and quantities, we place the actual bids (line 10), and a set of dummy bids in the market (line 11).

We first analyze if our proposed bidding strategy, MDPLCPBS, follows the Nash Equilibrium arrived in Section 3.2.2.1.1, and then benchmark it against the baseline and competing state-of-the-art strategies.

#### 3.2.3.4 Validation Experiments

We take the Power TAC simulator and isolate the wholesale market, and remove all market simulator participants (GenCos, internal buyers) from the market. We test the one buyer one seller (OBOS) scenario by deploying only two agents - a buyer and a seller. These agents participate only in the isolated wholesale market. They have a fixed energy demand they need to buy (sell) from the market. In these experiments, we set the energy demand to be the previous slot's tariff market net demand, which both the buyer and seller are notified about. We draw the valuations from a uniform distribution between 40 and 80, i.e. we set $l_S = l_B = 40$ and $h_S = h_B = 80$ in these experiments, and compute the theoretical scale factors using Equation (3.8) and Equation (3.9). We run two batches of experiments, with 30 games in each set of the batch (5 sets per batch). During each batch, one of the agents has a fixed scaling based bidding strategy, while the other uses MDPLCPBS.

In one batch, we draw the seller's valuation from the uniform distribution with bounds 40 and 80, and apply the theoretical scale factor (1.048689) to generate bids. Four other sets of experiments are also run, with scale factors within $\pm 0.1$ of the theoretical value. The buyer generates its valuation

| | Fixed seller's scale factor | | | | |
|---|---|---|---|---|---|
| Statistic / Scaling Factor | 0.948689 (-0.1) | 0.998689 (-0.05) | 1.048689 (Theoretical Value) | 1.098689 (+0.05) | 1.148689 (+0.1) |
| Average | 0.772435 | 0.804782 | 0.838087 | 0.863553 | 0.907389 |
| Standard Deviation | 0.033287 | 0.037697 | 0.025127 | 0.020749 | 0.036637 |

Table 3.1: Buyer's experimental scale factors values

| | Fixed buyer's scale factor | | | | |
|---|---|---|---|---|---|
| Statistic / Scaling Factor | 0.791386 (-0.1) | 0.841386 (-0.05) | 0.891386 (Theoretical Value) | 0.941386 (+0.05) | 0.991386 (+0.1) |
| Average | 0.989438 | 1.057427 | 1.113121 | 1.226423 | 1.616557 |
| Standard Deviation | 0.035598 | 0.027515 | 0.048204 | 0.036112 | 0.071974 |

Table 3.2: Seller's experimental scale factors values

from the uniform distribution with bounds 40 and 80, and uses this valuation as the $balancing\text{-}price$ in MDPLCPBS to generate bids.

In the second batch, we apply the buyer's theoretical scale factor (0.891386) to its true valuation (drawn from the same uniform distribution), to generate bids. Four other sets of experiments are also run, with scale factors within $\pm 0.1$ of the theoretical value. The seller, similar to the previous set of experiments, generates its true valuation and uses it as the $balancing\text{-}price$ in MDPLCPBS to bid in the market.

The experimental scale factor average and standard deviation for cleared bids for the buyer and the seller in the two batch of experiments are documented in Table 3.1 and Table 3.2. The values from Table 3.1 demonstrate that as the fixed scale factor for the seller is increased, the buyer's scale factor increases slowly. Table 3.2 demonstrates that as the fixed scale factor for the buyer is increased, the seller's scale factor increases rapidly. These tables demonstrate that in a one buyer and one seller setting, MDPLCPBS, while operating in a PDA, approaches the Nash Equilibrium stated in Section 3.2.2.1.1 for a single unit OBOS double auction with the selected parameter values.

#### 3.2.3.5 Benchmarks

We isolate the Power TAC wholesale market from the full Power TAC simulator while keeping the market simulator participants (GenCos, internal buyers) and weather simulator, and benchmark the performance of MDPLCPBS. The following agents/brokers are used in these benchmarks:

- **Zero Intelligence (ZI)**: The ZI agent [11] uses a randomized bid strategy and ignores the market state. They generate random order prices, ignoring the state of the market. For a given unit, prices are drawn from a uniform distribution between the unit's limit price and either a maximum

allowable price for sellers, or a minimum allowable price for buyers. In our experiments, we derive its bids from a uniform distribution with mean $\mu$ and a standard deviation of \$10. The mean $\mu$ taken from the limit price predicted by the MDP in TacTex [50]. The broker places one bid per auction, and the remaining required energy as the bid quantity. It continues to do the same for all the 24 bidding opportunities, or until the required energy is procured.

- **Zero Intelligence Plus (ZIP)**: The ZIP agent [48] maintains a scalar variable $m$ denoting its desired profit margin, and it combines this with a unit's limit price to compute a bid price $p$. For each failed trade, the price is adjusted by small increments to beat the failed bid price $p$. In our experiments, the initial limit price value $\mu$ is determined from the limit price predicted by the MDP in TacTex. The profit margin $m$ is set to $1\%$ of $\mu$, resulting in the initial bid price to be $p = \mu \times 1.01$. If the bid fails, the next bid price is incremented by $10\%$ of $\mu$. Then, the new bid price is given by $p = \mu \times 1.11$.
- **TacTex**: The TacTex [50] agent uses an MDP based model and dynamic programming to determine limit-prices for bids. The algorithm described in the paper was implemented and used in our experiments.
- **MCTS**: The MCTS [7] agent uses a Monte Carlo Tree Search (MCTS) coupled with heuristics on top of the limit price derived from a REPTree based limit price predictor, to determine the optimal bid price. In our experiments, we used the MCTS-dyn-C2 version with 10000 iterations, which is shown to be the best performing variation of the MCTS bidding strategy.

For a time-slot $t+24$ in the future, having 24 bidding opportunities in time-slots $\{t, t+1, \ldots, t+23\}$, the energy to be procured is set to be same across all the brokers. This energy amount for $t + 24$ is determined as some fraction of the net demand in time-slot $t$ in the Power TAC simulation tariff market. Four sets of 10 games each are simulated, with each set having a different fraction of the net demand to be procured. The fraction set is given by $\{0.25, 0.5, 0.75, 1\}$.

### 3.2.4 Experimental Analysis

Figure 3.2 shows the net cost of all the agents across the five sets of games. In each case, MDPLCPBS outperforms ZI, ZIP and TacTex on a consistent basis, while losing out to MCTS. It is to be noted that, while MCTS uses tailored heuristics, MDPLCPBS is derived from the game theoretic analysis of single shot double auction. We leave the game theoretic analysis of MCTS for future work.

## 3.3 Summary

In this chapter, we discussed solutions for a broker's two problems in the wholesale market - (i) total quantity to procure/sell from the market, and (ii) bidding policy to follow in the PDAs. To solve the first problem, we presented CUP, a neural network based usage predictor which uses the weather report, temporal data and past usage data to predict the customer's usage for a future time-slot. In order

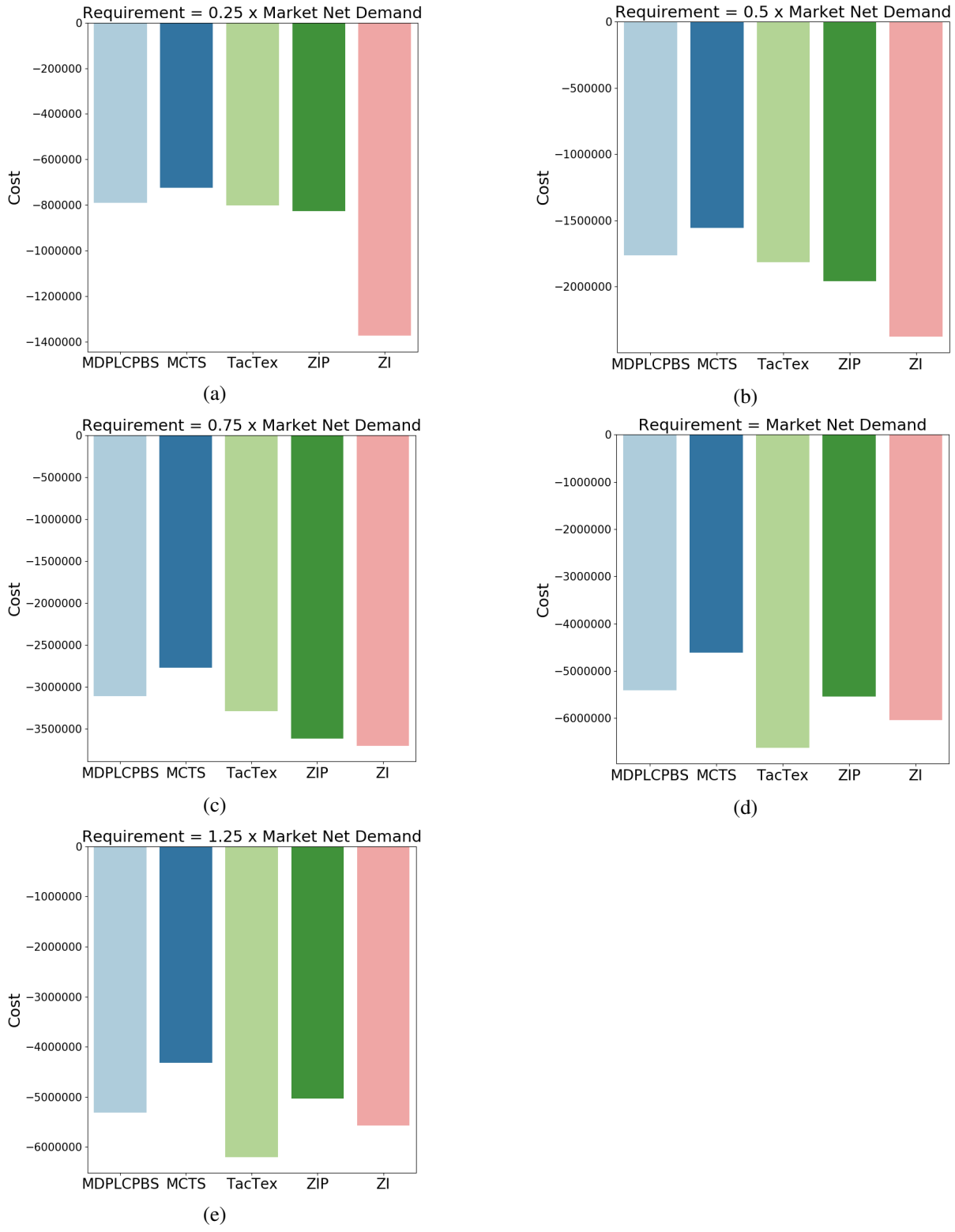


Figure 3.2: Net cost comparison of strategies across games with different energy requirements

to solve the second problem, we then presented MDPLCPBS, a bidding strategy for PDAs, derived from the game theoretic analysis of double auctions. In particular, we derived a Nash Equilibrium for a single unit double auction with the clearing price and payment rule as *ACPR*, for one buyer and one seller, and two buyers and one seller with scale based bidding strategies. We also derived the best response in a complete information setting in a multi-unit double auction with *ACPR*. Based on these formulations, we presented MDPLCPBS, a bidding strategy for PDAs. We experimentally showed that MDPLCPBS follows the Nash Equilibrium derived for single unit double auction with *ACPR*. We benchmarked MDPLCPBS against the baseline and competing state-of-the-art strategies, and showed that it outperforms most of them consistently.

*Chapter 4*

# Learning Strategies for the Tariff Market

This chapter focuses on the learning strategies for a broker in the tariff market. During a game, in order to attract customers and build a portfolio, a broker has to react to competing tariffs published in the market and publish tariffs accordingly. A broker can offer different types of tariffs (fixed, time-of-use, tiered, variable, etc.) during a game, any number of times, for different powertypes. The following sections discuss a reinforcement learning based strategy for a broker to devise time-of-use (TOU) tariffs, which try to minimize the impact of capacity charges while simultaneously trying to attract a large number of customers.

The tariff described in the following sections is an MDP based tariff, called the MDPTOU tariff. MDPTOU the result of solving an MDP using Q-learning, and is revised every twenty-four hours. Generating MDPTOU is a two-step process - (1) Generate a Fixed Price Tariff (FPT) by solving an MDP using Q-learning; (2) Convert the FPT to a TOU tariff for consumption customers by predicting the overall demand profile for the tariff market over the next 24 time slots. Section 4.1 describes the first process, while Section 4.2 and 4.3 discuss the latter.

## 4.1 MDP & Q-Learning Model (MDPQLM)

The following tariff MDP formulation is primarily motivated from the work of Cuevas et al. [9], and Reddy and Veloso [39]. The MDP augments the production tariff $P_{t,P}$ and consumption tariff $P_{t,C}$ of a learning broker, $B_L$, at time $t$. The MDP is defined as:

$$M^{B_L} = \langle S, A, \delta, R, \gamma \rangle \tag{4.1}$$

where $S$ is the state space, $A$ is the action space, $\delta(s, a)$ is the state transition function, $R(s, a)$ is the reward, and $\gamma$ is the discount factor. The elements of the MDP are defined as follows:

- **States**: At any simulation time $t$, the state of the MDP is a quadruple that captures four features of the tariff market. The set of states is described as:

$$S_t = \langle PRS_t, PS_t, CPS_t, PPS_t \rangle \tag{4.2}$$

where:

- $PRS_t$ = $\{rational, inverted\}$ is the price range status at time $t$
- $PS_t$ = $\{shortsupply, balanced, oversupply\}$ is the portfolio status at time $t$
- $CPS_t$ = $\{out, near, far, very-far\}$ is the consumer tariff price status at time $t$
- $PPS_t$ = $\{out, near, far, very-far\}$ is the producer tariff price status at time $t$

The first feature $PRS_t$ is rationality of the tariff market which is decided based on whether the highest production tariff is lower or higher than the lowest consumption tariff. The second is the portfolio status of the broker, $PS_t$, which could be surplus, balanced or deficit depending on the difference between the amount of energy acquired and committed in the tariff market at timet. They are defined as:

$$PRS_t = \begin{cases} rational, & \text{if } P_{t,C}^{\min} \geq P_{t,P}^{\max} \\ inverted, & \text{if } P_{t,C}^{\min} < P_{t,P}^{\max} \end{cases} \tag{4.3}$$

$$PS_t = \begin{cases} shortsupply, & \text{if } \theta_{t,T} < 0 \\ balanced, & \text{if } \theta_{t,T} = 0 \\ oversupply, & \text{if } \theta_{t,T} > 0 \end{cases} \tag{4.4}$$

where:

- $B_L$ is the learning broker using this strategy
- $P_{t,C}^{\min}$ = $\min_{B_k \in B \setminus \{B_L\}} P_{t,C}^{B_k}$ is the minimum active competing consumption price at time $t$ in the market.
- $P_{t,C}^{\max}$ = $\max_{B_k \in B \setminus \{B_L\}} P_{t,C}^{B_k}$ is the maximum active competing consumption price at time $t$ in the market.
- $P_{t,P}^{\min}$ = $\min_{B_k \in B \setminus \{B_L\}} P_{t,P}^{B_k}$ is the minimum active competing production price at time $t$ in the market.
- $P_{t,P}^{\max}$ = $\max_{B_k \in B \setminus \{B_L\}} P_{t,P}^{B_k}$ is the maximum active competing production price at time $t$ in the market.
- $\theta_{t,T}$ = $\theta_{t,P} - \theta_{t,C}$ is the net demand of the broker's customers at time $t$
- $\theta_{t,P}$: Total power sold in the tariff market by the broker's at time $t$
- $\theta_{t,C}$: Total power bought in the tariff market by the broker's at time $t$
- $B$: Set of all brokers participating in a game.

The third and fourth features, $PPS_t$ and $CPS_t$ respectively, rank the broker's current consumption and production tariffs with respect to prevailing tariffs of other competing broker agents, in a discrete manner. They are defined as:

$$CPS_t = \begin{cases} out, & \text{if } Top < P_{t,C} \\ near, & \text{if } Thres_{CPS} < P_{t,C} \leq Top \\ far, & \text{if } Middle < P_{t,C} \leq Thres_{CPS} \\ veryfar, & \text{if } P_{t,C} \leq Middle \end{cases} \tag{4.5}$$

$$PPS_t = \begin{cases} out, & \text{if } Bottom \geq P_{t,P} \\ near, & \text{if } Thres_{PPS} \geq P_{t,P} > Bottom \\ far, & \text{if } Middle \geq P_{t,P} > Thres_{PPS} \\ veryfar, & \text{if } P_{t,P} > Middle \end{cases} \tag{4.6}$$

where:

- $Top = P_{t,C}^{\min}$
- $Bottom = P_{t,P}^{\min}$
- $Middle = \frac{P_{t,C}^{\min} + P_{t,P}^{\min}}{2}$
- $Thres_{CPS} = \frac{Top + Middle}{2}$
- $Thres_{PPS} = \frac{Bottom + Middle}{2}$

In total, there are 96 possible states in the MDP. We use the *estimateCost* method provided by the Power TAC simulator to evaluate and normalize each tariff to a single number for the aforementioned comparison. It is to make sure that all possible combinations of tiered, TOU, dynamic and fixed tariffs are considered.

- **Actions**: A set of 8 actions are defined to augment the broker's consumption and production price over the course of the simulation. These actions allow the broker to suitably react to the changes in competing tariffs in the market, and are given as follows:
  - **Maintain**: Doesn't change the consumption or production prices. $P_{t+1,C}^{B_L} = P_{t,C}^{B_L}$, $P_{t+1,P}^{B_L} = P_{t,P}^{B_L}$
  - **Lower**: Decreases both prices by a fixed amount. $P_{t+1,C}^{B_L} = P_{t,C}^{B_L} - \delta_L$, $P_{t+1,P}^{B_L} = P_{t,P}^{B_L} - \delta_L$
  - **Raise**: Increases both prices by a fixed amount. $P_{t+1,C}^{B_L} = P_{t,C}^{B_L} + \delta_R$, $P_{t+1,P}^{B_L} = P_{t,P}^{B_L} + \delta_R$
  - **Revert**: Pushes both prices towards the midpoint $m_t$, defined as $m_t = \frac{P_{t,C}^{min} + P_{t,P}^{min}}{2}$
  - **Inline**: Set both prices near the midpoint. $P_{t+1,C}^{B_L} = m_t + \delta_{IL}$, $P_{t+1,P}^{B_L} = m_t - \delta_{IL}$

- **Wide**: Increases the consumption price and simultaneously reduces production price by fixed amount. $P^{B_L}_{t+1,C} = P^{B_L}_{t,C} + \delta_W$, $P^{B_L}_{t+1,P} = P^{B_L}_{t,P} - \delta_W$
- **MinMax**: $P^{B_L}_{t+1,C} = \alpha_M P^{max}_{t,C}$, $P^{B_L}_{t+1,P} = P^{min}_{t,P}$
- **Bottom**: $P^{B_L}_{t+1,C} = \alpha_B P^{min}_{t,C}$, $P^{B_L}_{t+1,P} = P^{min}_{t,P}$

where $\alpha_M \in [0.7, 1]$, and $\delta_L$, $\delta_R$, $\delta_{IL}$, $\delta_W$, $\alpha_M$ and $\alpha_B$ are all empirically optimized and determined.

- **Transition**: The transition function, $\delta$, is defined by numerous stochastic interactions between different customers and brokers within the Power TAC simulator. Defining $\delta$ being complex, as suggested in previous papers, we use reinforcement learning based approach to solve the MDP.

- **Reward**: The key novelty in our MDP formulation is the reward structure c.f. Cuevas et al.. The idea behind the reward structure is to capture the net profit made by the broker when it incurs no balancing fees. Thus, the reward at time $t$ is given by:

$$r_t = \theta_{t,C} P_{t,C} - \theta_{t,P} P_{t,P} - \theta_{t,W} W_t \tag{4.7}$$

The first term in Equation 4.7 represents the revenue generated by selling energy $\theta_{t,C}$ at the tariff $P_{t,C}$ to consumers of the broker at time $t$. Similarly, the second term represents the amount paid to producers of the broker for procuring energy $\theta_{t,P}$ at the tariff $P_{t,P}$. The third term in represents the amount paid in the wholesale market to satisfy the net unfulfilled demand $\theta_{t,W} = \theta_{t,C} - \theta_{t,P}$ at unit wholesale procurement cost $W_t$.

We construct a Q-table using Q-learning to solve the aforementioned MDP. For a state-action pair $(s_t, a_t)$, the Q-learning update rule with learning rate $\alpha$ and discount rate $\gamma$ is given by

$$\hat{Q}(s_t, a_t) \leftarrow (1 - \alpha_t)\hat{Q}(s_t, a_t) + \alpha[r_t + \gamma \max_a \hat{Q}(s_{t+1}, a)],$$

where $r_t$ is the reward obtained at time $t$ for taking action $a_t$ in state $s_t$. A Q-table is constructed through a training process in which this broker plays 100 games of every configuration against broker agents from past the Power TAC tournaments across different game configurations. For each configuration, the broker starts with a zero-initialized Q-table and updates the Q-table entries, across 100 games according to the update rule specified above. While playing a game to compete against other agents, at any tariff publication time $t$, being in state $s_t$, the broker is made to simply choose an action $a_t$ greedily according to $a_t = \text{argmax}_{a \in A} Q(s_t, a)$. The action thus chosen translates into a production FPT $P_{t,P}$ and a consumption FPT $P_{t,C}$. While the production FPT is published without any change, the consumption FPT is modified to generate MDPTOU as explained in *Tariff Designer* (TaD).

## 4.2 Net Demand Predictor (NDP)

Before converting the consumption FPT into MDPTOU, the broker needs to first estimate the overall net usage/demand of the tariff market for all future twenty-four time slots. To this end, at a simulation time slot $t$, the broker estimates the net demand $\hat{\mathrm{D}}_{t+k}, k \in \{1, \ldots, 24\}$ as a weighted average of two historical net demand values, namely, net demand $\mathrm{D}_{t+k-24}$ observed at the same time slot of the previous day and the net demand $\mathrm{D}_{t+k-168}$ observed during the same time-slot of the same day of the previous week. This enables the broker to capture the recent customer usage patterns in such a sensitive market while also utilizing the weekly trends. More specifically, we have,

$$\hat{\mathrm{D}}_{t+k} = \beta \mathrm{D}_{t+k-24} + (1-\beta)\mathrm{D}_{t+k-168} \tag{4.8}$$

where $\beta \in [0, 1]$ is a fixed parameter.

CUP is not used for this prediction task, as CUP may not have the per-customer usage data to predict the net demand of the market, as some customers are not subscribed to the broker and their usage data is not known to the broker. On the other hand, when the entire market demand is used as a usage statistic and fed into a separate neural network model in CUP, it performs poorly. This is because CUP, being trained on a per-customer level, fails to capture the combined effect of producers and consumers on a macro level. Hence we use a separate model, NDP, for predicting aggregate demand in the market. Also, on the contrary, NDP is not trained on per-customer level as the past 24 or 168 slot data may not available for new subscribers. Hence, NDPs prediction of the brokers net demand may not be good enough to place orders in wholesale market.

## 4.3 Tariff Designer (TaD)

Once the estimate of the net demand for the next twenty-four time slots is obtained from NDP, MDPTOU for a time slot $k$ hours ahead of the current time slot $t$ is computed as:

$$\pi_{t+k} = P_{t,C} + \rho \left( \hat{\mathrm{D}}_{t+k,T} - \frac{\sum_{j=1}^{24} \hat{\mathrm{D}}_{t+j,T}}{24} \right), \tag{4.9}$$

where $\rho$ is an empirically determined constant and $k \in \{1, \ldots, 24\}$. Equation (4.9) proposes MDPTOU for a twenty-four hour time horizon. Observe that the tariff rate in Equation (4.9) at a time slot $t$ modifies the fixed price consumption tariff $P_{t,C}$ provided by the Q-learning algorithm by an amount that is proportional to the excess estimated demand in that time slot over the mean estimated demand over the 24-hour period starting at $t$. The second term in Equation (4.9) closely resembles the manner in which the *transmission capacity fees* are calculated in the Power TAC simulation (see Section 2.1.2 of this thesis, or Section 7.2 of [19]). As a result, MDPTOU serves to mitigate the effect of transmission capacity fees that the broker incurs in two ways. First, it encourages the customers to shift some of their usage away from expected peak demand time-slot(s). Second, the excess over the consumption FPT charged to a customer is in proportion to that customer's contribution to the expected net demand

profile, and this helps offset some of the transmission capacity fees that will actually result from that customer's usage profile.

To avoid exponential explosion of the action space of the MDP in MDPQLM, we use a single value which is transformed into 24 hourly tariffs by TaD using Equation 4.9. This transformation ensures that the mean of the MDPTOU is the value chosen by the MDP, and this gets backpropagated to reflect the Q-values.Together, MDPQLM, NDP and TaD enable the broker to select actions in the large decision space of TOU tariffs by solving an MDP with a much smaller action space.

## 4.4 Summary

In this chapter, we described an MDP-based formulation of the tariff market, called the MDPQLM. While we took inspiration from Cuevas et al. and Reddy and Veloso for the MDP formulation, we modified the reward function, and solved the MDP using Q-learning. Furthermore, after arriving at the MDP solution, we described two heuristics which help in the transformation of fixed price tariffs (FPTs) into time-of-use (TOU) tariffs, namely the (1) Net Demand Predictor (NDP), and the (2) Tariff Designer (TaD). In the following chapter, we see these strategies perform in the Power TAC tournament, and analyze their individual performance using controlled offline experiments.

*Chapter 5*

# VidyutVanika: A Reinforcement Learning Based Broker Agent for a Power Trading Competition

This chapter introduces our autonomous broker agent, VidyutVanika, which participated in Power TAC 2018 Finals. VidyutVanika is a fully implemented broker agent, which implements the learning strategies described in Chapter 3 and 4 to bid intelligently in wholesale market, while simultaneously offering tariffs and reacting to tariffs in the tariff market. This chapter bridges the gap between theory and implementation (Section 5.1), describes the architecture of the VidyutVanika (Section 5.2), and analyzes VidyutVanika's successful performance in Power TAC 2018 Finals (Section 5.3). It also shows the efficacy of VidyutVanika's strategies and sub-modules in Section 5.4. VidyutVanika's binary is publicly available on the Power TAC broker binary repository (refer Related Publications & Releases)

## 5.1 From Theory to Practice

This section discusses the techniques associated to instantiate most the algorithms introduced in the previous chapters. While this section will discuss how hyper-parameters are tweaked for implementing some of these strategies, the actual values are not disclosed as these fine-tuned strategies may be used in future competitive scenarios.

The implementations were done using the sample broker code (provided by Power TAC) as skeleton, by replacing the appropriate modules with the corresponding strategies.

### 5.1.1 Wholesale Market Strategies

- **Customer Usage Predictor (CUP)**: The neural network implementation in CUP is done using Weka machine learning library for Java [13].

- **MDP and LCP based Bidding Strategy (MDPLCPBS)**: The MDPLCPBS is implemented as shown in Algorithm 1. We run a set of games to figure out the threshold for the amount of data points to be seen for $EnoughDataPoints()$ to be invoked. During these games, we test data point values for which the probability of clearance are significant for the available clearing prices.

During our tests, we found 24 data points to be a suitable value to start off, thus classifying it as the $EnoughDataPoints$ threshold.

Also, during the implementation, it is important to note that the action space of Limit Price Predictor (Section 3.2.3.1) is the real space. Since in practice it is difficult to work with the real space, we discretize it in the following manner. First, it is important to note that, given a set of past data points, i.e. LCPs and cleared quantities, the value of $p_{cleared}$ as computed in Equation 3.25 will be same between any two consecutive LCPs previously seen, when arranged in ascending order. This is because, for any of the values between two such LCPs, the cleared quantity remains the same. Thus, while implementing, we take the lower bound and add an $\epsilon$ margin and use it as our candidate limit price. We take the lower bound as we're trying the minimize over the limit-price in the dynamic programming value function. Thus, during our implementation, the action space is reduced to the number of distinct LCPs that the broker has observed till that time slot.

For implementing LCPP, we choose $\beta$ in such a way that it covers the possibility of estimating LCP when last cleared bid is lower than the limit-price that the broker bids. If the LCP ends up being higher, it is either way being estimated by any of the bids in between the range $[\beta \times limit\text{-}price, balancing\text{-}price]$. Thus, it is chosen to be closer to, but not less than 1.

### 5.1.2 Tariff Market Strategies

- **MDP & Q-Learning Model (MDPQLM):** The values of $\delta_L$, $\delta_R$, $\delta_{IL}$, and $\delta_W$ i.e. the increment/decrement values associated with the actions, are empirically determined. We choose these values in such a way that they do not skew the consumption and production prices too heavily, while also reflecting the varying usage patterns over the course of each day in the week. We determined these parameters in our implementation by analyzing the consecutive day-to-day usage patterns in the simulation, and adjusting these increments such that they can be operated across multiple weeks. If the tariffs in the market drop or increase suddenly, $\alpha_B$ and $\alpha_M$ readjust the prices in our implementation, with reference to the lowest and highest prices in the market. They are thus chosen to be between 0.7 and 1, after analyzing such scenarios across multiple simulations, in order to remain competitive in the tariff market, and respond to abrupt tariff swings.

  We also establish a baseline tariff threshold in our implementation, below which our consumption tariff is not reduced. This is determined using the wholesale market price and small profit margin above it. Anytime the MDP takes an action which decreases the consumption tariff below such baseline threshold, the consumption tariff is set to this threshold. Similarly we have an upper threshold for production tariffs (value beyond which production tariff is not incremented) which is calculated in a similar way.

  During our training process to construct the Q-table, our implemented broker played 100 games of 8 different configurations against broker agents from past the Power TAC tournaments. For each configuration, the broker started with a zero-initialized Q-table, updated the Q-table entries,

and saved them. Training was done across the games, while keeping per-configuration Q-tables separate. While testing, the broker was tasked to load the appropriate Q-table depending on the game configuration i.e. number of brokers participating in the game.

- **Tariff Designer (TaD):** While implementing the TaD, the $\rho$ in Equation 4.9 was determined in such a way that the twenty-four hour rates were mostly within a range of $\pm 0.1$. Further, after determining the rates, they were checked to have satisfied the per-time slot baseline tariff threshold, similar to the implementation of MDPQLM rates. If a rate came out to be below the threshold, it was set to the threshold value. This was done to ensure that the implemented strategy was not giving out loss-making irrational rates for certain time slots.

## 5.2 VidyutVanika: Architecture and Strategy

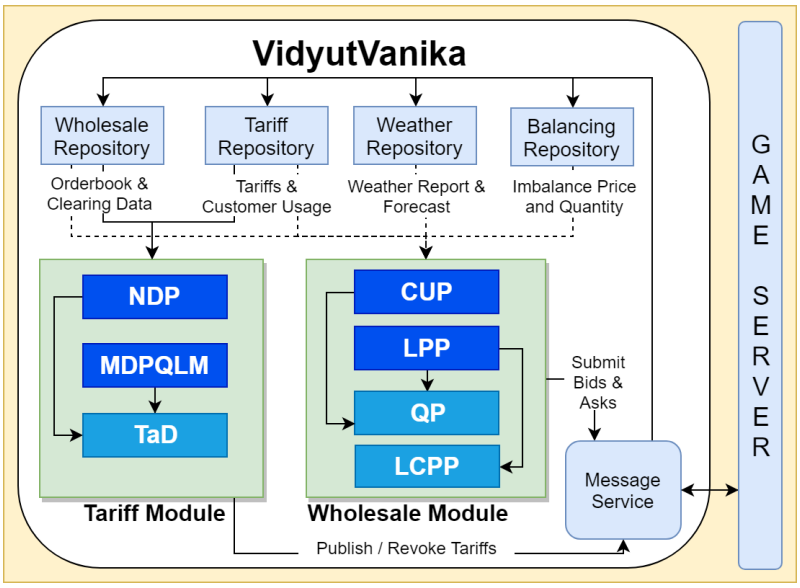


Figure 5.1: Architecture of VidyutVanika

VidyutVanika, abbreviated as $VV$, is an amalgamation of the words 'Vidyut' which means electricity, and 'Vanika' which means broker. VidyutVanika incorporates the learning strategies of the wholesale market as described in Chapter 3, and that of the tariff market as described Chapter 4. VidyutVanika is written in Java, and uses JMS to communicate with the game server. Figure 5.1 describes the architecture of VidyutVanika.

VidyutVanika consists of two main modules, namely, Tariff Module (TM) and Wholesale Module (WM). TM is responsible for publishing and revoking tariffs in the tariff (or retail) market. The tariff market strategies, namely NDP, MDPQLM and TaD, form the TM. During a game, VidyutVanika offers two active time-of-use tariffs - (i) MDPTOU (described in Chapter 4), and (ii) WeeklyTOU. WeeklyTOU is an empirically determined, fixed, weekly TOU tariff, which remains active throughout the duration of

the game. If MDPTOU makes losses for a sustained period of time, VidyutVanika revokes it and falls back upon WeeklyTOU, as it is empirically proven to be reliable. On the other hand, WM generates bids/asks to purchase/sell energy contracts in the wholesale market. The wholesale market strategies of CUP, LPP, QP and LCPP, as discussed in Chapter 3 form the WM. VidyutVanika doesnt actively participate in the balancing market. We note that, to the best of our knowledge, VidyutVanika is the first broker agent to use neural networks with the weather data to predict customer usage in Power TAC competition.

## 5.3 Power TAC 2018 Finals Results

We analyze the performance of our broker VidyutVanika in Power TAC 2018 Finals. The Power TAC 2018 Finals had 7 brokers from research groups across the globe. The tournament had a total of 324 games, with all possible combinations of 7-broker games (100 games), 4-broker games (140 games; 80 games for each broker), and 2-broker games (84 games; 24 games for each broker). Table 5.1 shows the net profit of all brokers across different game configurations, percentage of profit in comparison to the winning agent, AgentUDE, and the corresponding normalized scores. Despite winning more games than AgentUDE, VidyutVanika was placed next to AgentUDE in overall ranking of Power TAC 2018. This is because, the determination of the winner is made based on normalized cumulative profits in each configuration across all games in the tournament. Specifically, AgentUDE netted high profits against competing agents (excluding VidyutVanika) in 2-player games that helped in cementing its place as the winner of the tournament.

| Broker | 7-broker | 4-broker | 2-broker | Total | 7-broker (N) | 4-broker (N) | 2-broker (N) | Total (N) |
|---|---|---|---|---|---|---|---|---|
| AgentUDE | 49964603 (100) | 62138484 (100) | 134908672 (100) | 247011760 (100) | 1.091 | 0.634 | 1.565 | 3.291 |
| VidyutVanika | 48197051 (96) | 101942819 (164) | 47541635 (35) | 197681504 (80) | 1.056 | 1.061 | 0.336 | 2.453 |
| CrocodileAgent | 27659543 (55) | 45441732 (73) | 62881837 (47) | 135983111 (55) | 0.648 | 0.455 | 0.552 | 1.655 |
| SPOT | -6979768 (-14) | 32981756 (53) | 49183707 (36) | 75185695 (30) | -0.041 | 0.322 | 0.359 | 0.64 |
| COLDPower18 | 2063729 (4) | 10289982 (17) | 521330 (0.3) | 12875040 (5) | 0.139 | 0.078 | -0.326 | -0.109 |
| Bunnie | -67983216 (-136) | -25049555 (-40) | -19596577 (-15) | -112629348 (-46) | -1.254 | -0.3 | -0.609 | -2.163 |
| EWIIS3 | -87271195 (-175) | -206960249 (-333) | -109800161 (-81) | -404031605 (-164) | -1.638 | -2.25 | -1.878 | -5.766 |

Table 5.1: Power TAC 2018 – Net profits and normalized scores (denoted by (N)) of each broker

Table 5.2 shows the number of $1^{st}$ and $2^{nd}$ place finishes by each broker across all three configurations. As seen, VidyutVanika won the most number of games in the tournament with 112 wins out of the 204 it participated in, with AgentUDE coming second with 92 wins out of 204. VidyutVanika had the most wins in 7-broker and 4-broker games, and had the second highest number of wins, behind AgentUDE, in 2-broker games. It is important to note that, overall, VidyutVanika finished in the top two, 72% of the time whenever it played in a game with more than 2 brokers. In comparison, AgentUDE stood at 65%. On a head-to-head comparison with AgentUDE, out of 100 7-broker games, AgentUDE and VidyutVanika both shared 39 wins each. However in 4-Broker games in which both VidyutVanika and AgentUDE participated, VidyutVanika won 31 times out 40, with AgentUDE winning the remaining 9. In the four 2-broker games involving both brokers, AgentUDE ended up winning three games. Vidyut-

Vanika led in all these three lost games almost till the end, only to fall behind finally due to transmission capacity fees. Figure 5.2 shows the number of games in which each broker ended up with a negative profit. CrocodileAgent had the fewest games with negative profits, with VidyutVanika coming second in this category with four times the average market share. Thus, VidyutVanika managed to make up for its losses on a consistent basis, and rarely ended up being non-profitable.

| Brokers | 7-Broker | | 4-Broker | | 2-Broker | | Total | |
|---|---|---|---|---|---|---|---|---|
| | $1^{st}$ | $2^{nd}$ | $1^{st}$ | $2^{nd}$ | $1^{st}$ | $2^{nd}$ | $1^{st}$ | $2^{nd}$ |
| VidyutVanika | 39 | 21 | 54 | 14 | 19 | 5 | 55 | 20 |
| AgentUDE | 39 | 26 | 31 | 21 | 22 | 2 | 45 | 24 |
| CrocodileAgent | 8 | 34 | 13 | 41 | 15 | 9 | 18 | 41 |
| SPOT | 0 | 0 | 16 | 19 | 9 | 15 | 12 | 17 |
| COLDPower18 | 0 | 3 | 5 | 29 | 8 | 16 | 6 | 24 |
| Bunnie | 13 | 15 | 21 | 16 | 9 | 15 | 21 | 22 |
| EWIIS3 | 1 | 1 | 0 | 0 | 2 | 22 | 1 | 11 |

Table 5.2: Power Tac 2018 – Number of $1^{st}$ and $2^{nd}$ place standings of each broker

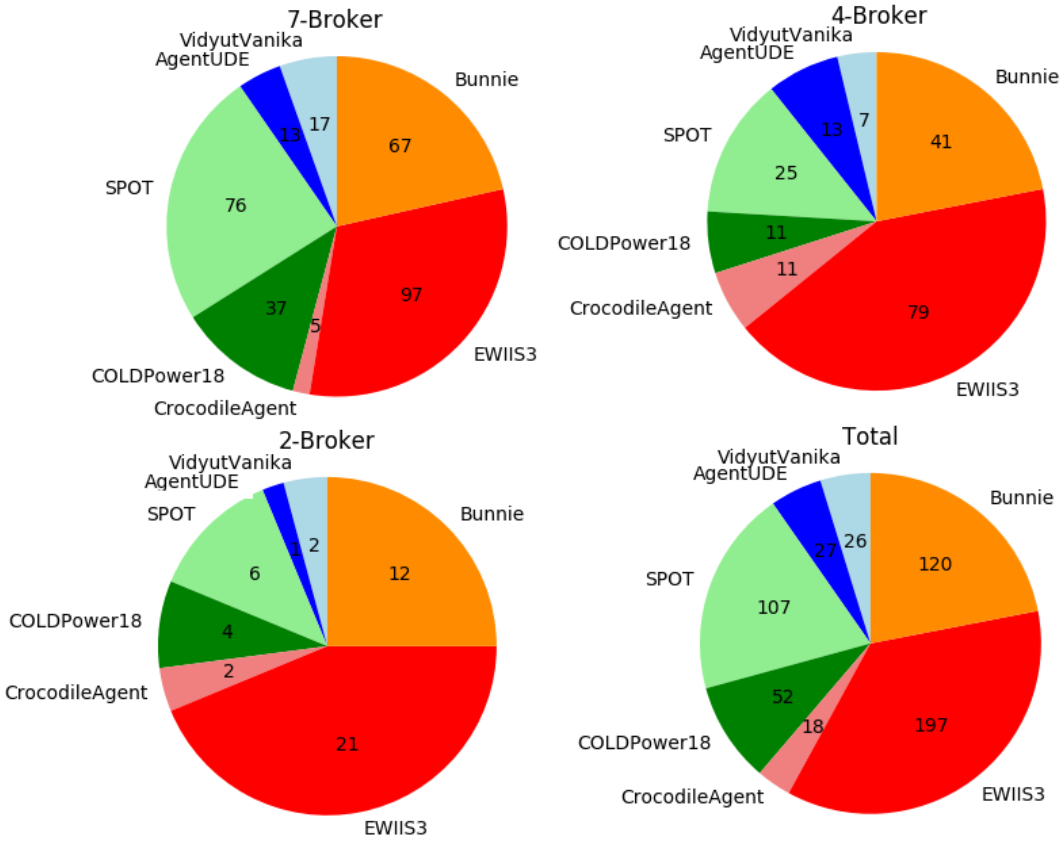


Figure 5.2: Power TAC 2018 – Number of games with negative profits

TM played a crucial role in VidyutVanika's success, offering tariffs which were attractive to majority of the customers and contributed the most in revenue. Figure 5.3 shows the average market share to each broker across all three configurations and overall. Note that the percentage will not sum up to 100 in some configurations. E.g.: In 4-broker games, each broker plays 80 games, where as in total 140 games are played. VidyutVanika had the highest market share on average in 2-broker games, 7-broker games and overall, and the second highest in 4-broker games. In contrast, AgentUDE had only a quarter of the overall average market share of VidyutVanika. While one may expect a greater market share to lead to more profits, it usually leads to higher transmission capacity fees and distribution costs, which can cause higher losses unless managed properly. As a result, agents with lower market share often tend to make less losses, and end up winning. Figure 5.4 represents the average income and costs of all brokers across all three configurations. VidyutVanika clearly has less imbalance costs while having almost similar number of customers as Bunnie, exhibiting the effectiveness of CUP. VidyutVanika also had one of the best tariff market income-to-cost ratio (1.14), with only AgentUDE (1.43) and CrocodileAgent (1.32) having better ratios. However, both AgentUDE and CrocodileAgent had very low average market share compared to VidyutVanika. Thus, VidyutVanika is very efficient at making profits despite having a higher market share.

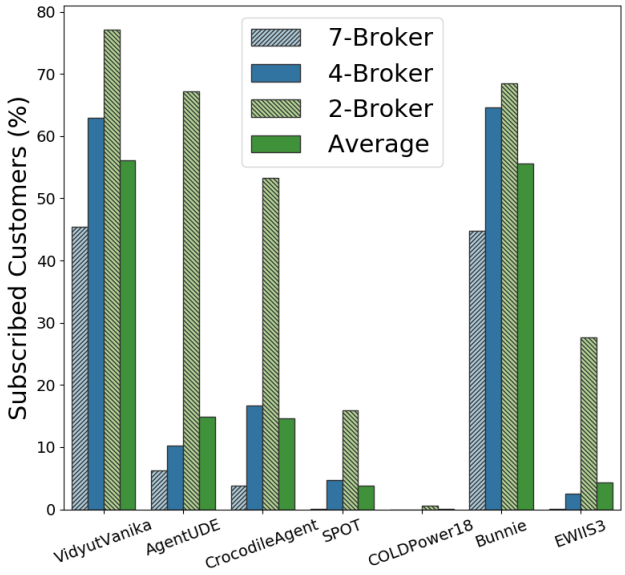


Figure 5.3: Power TAC 2018 – Average Percentage of customers subscribed (out of 57000), i.e. market share, of each broker

## 5.4 Controlled Offline experiments

For all controlled offline experiments, we played games using randomly chosen weather files from the 324 games played in the Power TAC 2018 Finals.

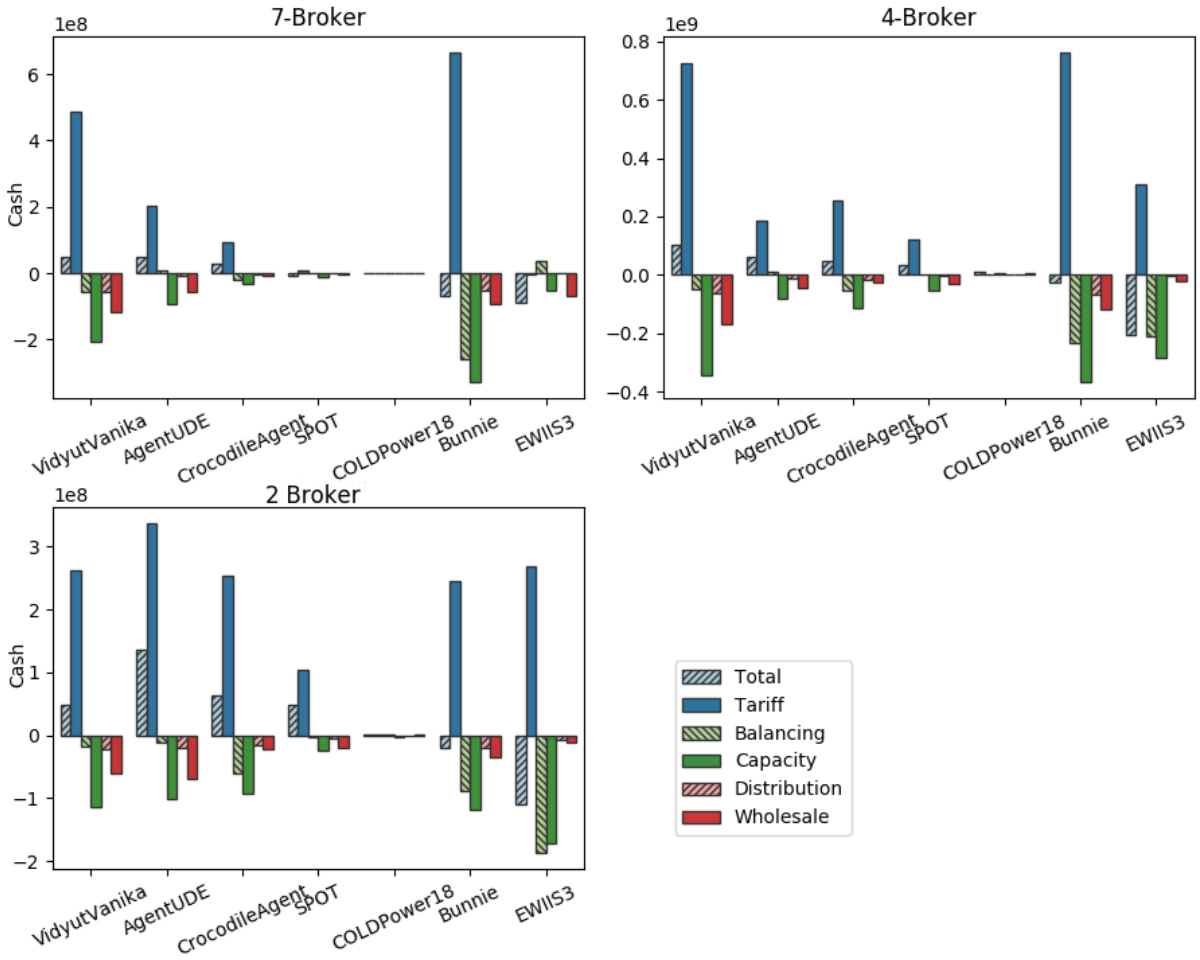


Figure 5.4: Power TAC 2018 – Average Income/Costs of each broker

To determine the prediction accuracy of NNs used in Customer Usage Predictor (CUP), we played a set of 30 games with VidyutVanika being the sole participant. As all the customers end up subscribed to VidyutVanika in such a game, we determined the accuracy of each customer's usage prediction by comparing it to their actual usage. We got an average prediction accuracy of 84% from these set of games.

Next, in order to identify the contribution of each submodule in VidyutVanika, we performed controlled offline experiments with test agents created by disabling multiple combinations of submodules. Agent F_LCPP is generated from VidyutVanika by disabling the Last Cleared Price Predictor (LCPP) submodule, and replacing the LCP in the LPP MDP solution with the clearing price, as implemented by [50]. Agent F_QP is generated from VidyutVanika by disabling Quantity Predictor (QP) submodule, which results in the agent placing the entire predicted net demand in a single bid in the wholesale market. Agent F_LCPP_QP is generated from VidyutVanika by disabling both the LCPP and QP submodules as above. Agent F_CUP is generated from VidyutVanika by replacing CUP by the usage predictor provided in the Power TAC *sample broker* which essentially predicts customer usage by exponentially smoothing

over the past usage records, incrementally. Agent F_WM is generated by disabling the entire Wholesale Module, and replacing it by the wholesale strategy provided in the Power TAC *sample broker*. F_WM essentially increases the limit prices as the target time-slot gets closer with some randomization in the limit price determination. Agent F_Reward is generated from VidyutVanika by replacing the MDPQLM reward function by the reward function used by Cuevas et al.. Agent F_TaD is generated from VidyutVanika by disabling Tariff Designer (TaD) and instead offering FPTs from MDPQLM. In theory, this agent has the same tariff strategy as proposed by [9], but with our reward function. Agent F_TaD_CUP is generated by disabling both TaD and CUP as described above. Agent F_TM is generated from VidyutVanika by disabling TM, but keeping WeeklyTOU active. Finally, F_TM_WM is generated by disabling both TM and WM in the manner described above.

| Brokers | % of VidyutVanika's profit |
|---|---|
| F_Reward | 75 |
| F_TaD | 83 |
| F_CUP | 84 |
| F_TaD_CUP | 75 |
| F_TM | 73 |
| F_LCPP | 76 |
| F_QP | 79 |
| F_LCPP_QP | 71 |
| F_WM | 90 |
| F_TM_WM | 72 |

Table 5.3: Performance of Test Agents vs the full agent VidyutVanika

Each of these test agents were made to compete with the full agent VidyutVanika over 30 games. The results of these experiments are reported in Table 5.3. Both TM and WM offer significant improvements as compared to the base sample broker strategy, with the former playing the biggest part in VidyutVanika's success. CUP, LCPP and QP submodules play a crucial role in VidyutVanika's wholesale market strategy, and cause a significant decrease in profit when removed, as seen from the table. On the other hand, TaD submodule (responsible for generating MDPTOU) is crucial to VidyutVanika's tariff market strategy, removal of which causes a sharp decline in the broker's profit. Also note that, there is a significant decrease in profit when we used the reward function from [9][1] in F_Reward.

## 5.5 Summary

In this chapter, we bridged the gap between the strategies developed, and their implementations. We described our autonomous energy broker, VidyutVanika, and its architecture. VidyutVanika implemented our learning strategies described in Chapter 3 and 4, along with a few heuristics. VidyutVanika finished as the runner-up in the Power TAC 2018 Finals. We analyzed VidyutVanika's performance in

[1] We used a suitable value for the hyper-parameter in their reward function

the Power TAC 2018 Finals, and the reasons behind its success. Furthermore, we showed the efficacy of our learning strategies and heuristics using controlled offline experiments.

*Chapter 6*

# Conclusion

We described our learning strategies for the wholesale and tariff markets of smart grids, using the Power TAC simulator as our application domain for testing. We presented MDPLCPBS, a bidding strategy for Periodic Double Auctions, derived from the game theoretic analysis of double auctions. In particular, we derived a Nash Equilibrium for a single unit double auction with the clearing price and payment rule as ACPR, for one buyer and one seller, and two buyers and one seller with scale based bidding strategies. We also derived the best response in a complete information setting in a multi-unit double auction with ACPR. Based on these formulations, we presented MDPLCPBS, a bidding strategy for PDAs, which applies to Power TAC wholesale market PDAs. We experimentally showed that MDPLCPBS follows the Nash Equilibrium derived for single unit double auction with ACPR. We benchmarked MDPLCPBS against the baseline and competing state-of-the-art strategies, and showed that it outperforms most of them consistently. We also presented a neural network based usage predictor, named CUP, to predict future demands of customers. It was the first of its kind to use weather data and forecast in the Power TAC setting.

We then presented the formulation of MDP for the tariff markets in smart grid. The MDP reward function, heuristics for solution transformation, and application was the another key contribution of our work. Furthermore, we described our own autonomous energy broker, VidyutVanika, and its architecture. VidyutVanika was the runner-up in Power TAC 2018 Finals. We illustrated how our learning strategies for both tariff and wholesale markets formed the core of VidyutVanika. We proceeded to analyze VidyutVanika's performance and also illustrated the efficacy of our strategies by providing the detailed analysis of: (i) the comparative market-wise performance of VidyutVanika in the Power TAC 2018 Finals and (ii) the offline experiments to demonstrate the contribution of each sub-module of VidyutVanika, which instantiate our strategies and heuristics.

# Related Publications & Releases

## Conference Publications

1. **Ghosh, S.**, Subramanian, E., Bhat, S.P., Gujar, S. and Paruchuri, P., 2019. *VidyutVanika: A Reinforcement Learning Based Broker Agent for a Power Trading Competition*. To be **published** in the proceedings of Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19) held in Honolulu, Hawaii, USA from January 27 - February 1 2019.

2. **Ghosh, S.**, Prakash K., Chandlekar S., Subramanian E., Bhat, S.P., Gujar, S. and Paruchuri, P., 2019. *VidyutVanika: An Autonmous Broker Agent for Smart Grid Environment*. **Accepted** for Policy, Awareness, Sustainability and Systems (PASS) Workshop, to be held in Cologne, Germany from June 25 - 26, 2019.

## Releases

1. **VidyutVanika 2018** broker binary, published on the Power TAC broker repository. URL - `http://www.powertac.org/wiki/index.php/VidyutVanika`. Alternate URL - `http://researchweb.iiit.ac.in/~susobhan.ghosh/VidyutVanika.zip`

# Bibliography


[1] ABB. Solutions for Smart Grid. `http://new.abb.com/smartgrids/what-is-a-smart-grid`. [Online; accessed 20-May-2019].

[2] J. Babic and V. Podobnik. Adaptive bidding for electricity wholesale markets in a smart grid. In *AAMAS Workshop on Agent-Mediated Electronic Commerce and Trading Agents Design and Analysis (AMEC/TADA 2014)*, 2014.

[3] R. Bellman. *Dynamic programming*. Courier Corporation, 2013.

[4] K. Chatterjee and W. Samuelson. Bargaining under incomplete information. *Operations research*, 31(5):835–851, 1983.

[5] M. M. P. Chowdhury. Predicting prices in the Power TAC wholesale energy market. In *AAAI*, pages 4204–4205, 2016.

[6] M. M. P. Chowdhury, R. Y. Folk, F. Fioretto, C. Kiekintveld, and W. Yeoh. Investigation of learning strategies for the SPOT broker in Power TAC. In S. Ceppi, E. David, C. Hajaj, V. Robu, and I. A. Vetsikas, editors, *Agent-Mediated Electronic Commerce. Designing Trading Strategies and Mechanisms for Electronic Markets*, pages 96–111, Cham, 2017. Springer International Publishing. ISBN 978-3-319-54229-4.

[7] M. M. P. Chowdhury, C. Kiekintveld, T. C. Son, and W. Yeoh. Bidding strategy for periodic double auctions using Monte Carlo tree search. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems*, pages 1897–1899. International Foundation for Autonomous Agents and Multiagent Systems, 2018.

[8] E. Commission. *Energy roadmap 2050*. Publications Office of the European Union, 2012.

[9] J. S. Cuevas, A. Y. Rodriguez-Gonzalez, and E. M. De Cote. Fixed-price tariff generation using reinforcement learning. In *Modern Approaches to Agent-based Complex Automated Negotiation*, pages 121–136. Springer, 2017.

[10] R. Eberhart and J. Kennedy. A new optimizer using particle swarm theory. In *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, pages 39–43. Ieee, 1995.

[11] D. K. Gode and S. Sunder. Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of political economy*, 101(1):119–137, 1993.

[12] D. Grgic, H. Vdovic, J. Babic, and V. Podobnik. Crocodileagent 2018: Robust agent-based mechanisms for power trading in competitive environments. *Comput. Sci. Inf. Syst.*, 16(1):105–129, 2019.

[13] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[14] J. Hoogland and H. La Poutré. An effective broker for the power tac 2014. In *Agent-Mediated Electronic Commerce. Designing Trading Strategies and Mechanisms for Electronic Markets*, pages 66–80. Springer, 2015.

[15] W. Ketter, J. Collins, and C. A. Block. Smart grid economics: Policy guidance through competitive simulation. 2010.

[16] W. Ketter, J. Collins, and P. Reddy. Power tac: A competitive economic simulation of the smart grid. *Energy Economics*, 39:262–270, 2013.

[17] W. Ketter, M. Peters, J. Collins, and A. Gupta. Competitive benchmarking: an is research approach to address wicked problems with big data and analytics. 2015.

[18] W. Ketter, M. Peters, J. Collins, and A. Gupta. A multiagent competitive gaming platform to address societal challenges. *Mis Quarterly*, 40(2):447–460, 2016.

[19] W. Ketter, J. Collins, and M. Weerdt. The 2018 power trading agent competition, 2017. URL `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3087096/`.

[20] R. T. Kuate, M. He, M. Chli, and H. H. Wang. An intelligent broker agent for energy trading: an mdp approach. In *Twenty-Third International Joint Conference on Artificial Intelligence*, 2013.

[21] R. T. Kuate, M. Chli, and H. H. Wang. Optimising market share and profit margin: Smdp-based tariff pricing under the smart grid paradigm. In *IEEE PES Innovative Smart Grid Technologies, Europe*, pages 1–6. IEEE, 2014.

[22] B. Liefers, J. Hoogland, and H. La Poutré. A successful broker agent for Power TAC. In *Agent-Mediated Electronic Commerce. Designing Trading Strategies and Mechanisms for Electronic Markets*, pages 99–113. Springer, 2014.

[23] S. Matetic, J. Babic, M. Matijas, A. Petric, and V. Podobnik. The crocodileagent 2012: Negotiating agreements in smart grid tariff market. In *AT*, pages 203–204, 2012.

[24] Ministry of Power, Government of India. Smart Grid Vision and Roadmap for India. http://www.indiasmartgrid.org/reports/Smart%20Grid%20Vision%20and%20Roadmap%20for%20India.pdf, 2019. [Online; accessed 20-May-2019].

[25] Ministry of Science & Technology, Government of India. India Country Report on Smart Grids. http://dst.gov.in/sites/default/files/India%20Country%20Report%20on%20Smart%20Grids.pdf, 2017. [Online; accessed 20-May-2019].

[26] N. Nakamura. Preparing for Future Energy Demands with Smart Grid Technology. https://electricenergyonline.com/energy/magazine/1163/article/Preparing-for-Future-Energy-Demands-with-Smart-Grid-Technology.htm, 2019. [Online; accessed 20-May-2019].

[27] Y. Narahari. *Game theory and mechanism design*, volume 4. World Scientific, 2014.

[28] Y. Narahari. *Game theory and mechanism design*, volume 4, pages 197–198. World Scientific, 2014.

[29] Nord Pool AS. Nord Pool key statistics - March 2018. https://www.nordpoolgroup.com/message-center-container/newsroom/exchange-message-list/2018/q2/nord-pool-key-statistics--march-2018/, 2018. [Online; accessed 20-May-2019].

[30] E. Ntagka, A. Chrysopoulos, and P. A. Mitkas. Designing tariffs in a competitive energy market using particle swarm optimization techniques. In *Agent-Mediated Electronic Commerce. Designing Trading Strategies and Mechanisms for Electronic Markets*, pages 129–143. Springer, 2014.

[31] S. Özdemir and R. Unland. Autonomous power trading approaches of a winner broker. In *AAMAS Workshop on Agent-Mediated Electronic Commerce and Trading Agents Design and Analysis (AMEC/TADA 2015)*, pages 143–156. Springer, 2015.

[32] S. Özdemir and R. Unland. AgentUDE17: Imbalance management of a retailer agent to exploit balancing market incentives in a smart grid ecosystem, 03 2018. URL http://mkwi2018.leuphana.de/wp-content/uploads/MKWI_319.pdf.

[33] S. Özdemir and R. Unland. AgentUDE17: A genetic algorithm to optimize the parameters of an electricity tariff in a smart grid environment. In *Advances in Practical Applications of Agents, Multi-Agent Systems, and Complexity: The PAAMS Collection*, pages 224–236. Springer, 2018.

[34] S. Parsons, J. A. Rodriguez-Aguilar, and M. Klein. Auctions and bidding: A guide for computer scientists. *ACM Computing Surveys (CSUR)*, 43(2):10, 2011.

[35] C. Perlich, B. Dalessandro, R. Hook, O. Stitelman, T. Raeder, and F. Provost. Bid optimizing and inventory scoring in targeted online advertising. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 804–812. ACM, 2012.

[36] M. Peters, W. Ketter, M. Saar-Tsechansky, and J. Collins. Autonomous data-driven decision-making in smart electricity markets. In P. A. Flach, T. De Bie, and N. Cristianini, editors, *Machine Learning and Knowledge Discovery in Databases*, pages 132–147, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[37] M. Peters, W. Ketter, M. Saar-Tsechansky, and J. Collins. A reinforcement learning approach to autonomous decision-making in smart electricity markets. *Machine Learning*, 92(1):5–39, 7 2013. ISSN 0885-6125. doi: 10.1007/s10994-013-5340-0.

[38] M. L. Puterman. Markov decision processes. *Wiley and Sons*, 1994.

[39] P. P. Reddy and M. M. Veloso. Strategy learning for autonomous agents in smart grid markets. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Two*, IJCAI'11, pages 1446–1451. AAAI Press, 2011. ISBN 978-1-57735-514-4. doi: 10.5591/978-1-57735-516-8/IJCAI11-244. URL http://dx.doi.org/10.5591/978-1-57735-516-8/IJCAI11-244.

[40] P. P. Reddy and M. M. Veloso. Negotiated Learning for Smart Grid Agents: Entity Selection Based on Dynamic. In *Proceedings of AAAI'13, the Twenty-Seventh AAAI Conference on Artificial Intelligence*, Bellevue, Washington, July 2013.

[41] M. H. Rothkopf. Equilibrium linear bidding strategies. *Operations Research*, 28(3-part-i):576–583, 1980.

[42] T. R. Rúbio, J. Queiroz, H. L. Cardoso, A. P. Rocha, and E. Oliveira. TugaTAC broker: A fuzzy logic adaptive reasoning agent for energy trading. In *Multi-Agent Systems and Agreement Technologies*, pages 188–202. Springer, 2015.

[43] S. J. Russell and P. Norvig. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,, 2016.

[44] B. Speer, M. Miller, W. Schaffer, L. Gueran, A. Reuter, B. Jang, and K. Widegren. Role of smart grids in integrating renewable energy. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2015.

[45] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[46] R. S. Sutton, D. Precup, and S. Singh. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1-2):181–211, 1999.

[47] G. Tesauro and J. L. Bredin. Strategic sequential bidding in auctions using dynamic programming. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 2*, AAMAS '02, pages 591–598, New York, NY, USA, 2002. ACM. ISBN 1-58113-480-0. doi: 10.1145/544862.544885. URL http://doi.acm.org/10.1145/544862.544885.

[48] G. Tesauro and R. Das. High-performance bidding agents for the continuous double auction. In *Proceedings of the 3rd ACM conference on Electronic Commerce*, pages 206–209. ACM, 2001.

[49] T. Urban and W. Conen. Maxon16: A successful Power TAC broker. In *International Workshop on Agent-Mediated Electronic Commerce and Trading Agents Design and Analysis (AMEC/TADA 2017)*, 2017. URL `http://www.sofiaceppi.com/AMECTADA2017/AMEC_TADA-17.pdf`.

[50] D. Urieli and P. Stone. TacTex'13: A champion adaptive power trading agent. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pages 465–471. Association for the Advancement of Artificial Intelligence, 2014.

[51] D. Urieli and P. Stone. An MDP-based winning approach to autonomous power trading: formalization and empirical analysis. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 827–835. International Foundation for Autonomous Agents and Multiagent Systems, 2016.

[52] D. Urieli and P. Stone. Autonomous electricity trading using time-of-use tariffs in a competitive market. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*. Association for the Advancement of Artificial Intelligence, 2016.

[53] US Department of Energy. Grid 2030. *A National Vision for Electricity's Second*, 100, 2003.

[54] US Department of Energy. Smart Grid System Report. `https://www.energy.gov/sites/prod/files/2019/02/f59/Smart%20Grid%20System%20Report%20November%202018_1.pdf`, 2018. [Online; accessed 20-May-2019].

[55] D. R. Vincent. Bidding off the wall: Why reserve prices may be kept secret. *Journal of Economic theory*, 65(2):575–584, 1995.

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

[57] R. Wilson. Strategic analysis of auctions. *Handbook of game theory with economic applications*, 1:227–279, 1992.

[58] P. R. Wurman, W. E. Walsh, and M. P. Wellman. Flexible double auctions for electronic commerce: Theory and implementation. *Decision Support Systems*, 24(1):17–27, 1998.

[59] Y. Yang, J. Hao, M. Sun, Z. Wang, C. Fan, and G. Strbac. Recurrent deep multiagent Q-learning for autonomous brokers in smart grid. In *IJCAI*, pages 569–575, 2018.