An MDP-Based Winning Approach to Autonomous Power Trading: Formalization and Empirical Analysis

Daniel Urieli
Dept. of Computer Science
University of Texas at Austin
Austin, TX 78712, USA
urieli@cs.utexas.edu

Peter Stone
Dept. of Computer Science
University of Texas at Austin
Austin, TX 78712, USA
pstone@cs.utexas.edu

ABSTRACT

With the efforts of moving to sustainable and reliable energy supply, electricity markets are undergoing far-reaching changes. Due to the high-cost of failure in the real-world, it is important to test new market structures in simulation. This is the focus of the Power Trading Agent Competition (Power TAC), which proposes autonomous electricity broker agents as a means for stabilizing the electricity grid. This paper focuses on the question: how should an autonomous electricity broker agent act in competitive electricity markets to maximize its profit. We formalize the electricity trading problem as a continuous, high-dimensional Markov Decision Process (MDP), which is computationally intractable to solve. Our formalization provides a guideline for approximating the MDP's solution, and for extending existing solutions. We show that a previously champion broker can be viewed as approximating the solution using a lookahead policy. We present TacTex'15, which improves upon this previous approximation and achieves state-of-the-art performance in competitions and controlled experiments. Using thousands of experiments against 2015 finalist brokers, we analyze TacTex'15's performance and the reasons for its success. We find that lookahead policies can be effective, but their performance can be sensitive to errors in the transition function prediction, specifically demand-prediction.

Keywords

Autonomous electricity trading; Trading Agents; Smart-grid; Markets

1. INTRODUCTION

With the efforts of moving to sustainable and reliable energy supply, electricity markets (aka power markets) are undergoing far-reaching changes: customers are being engaged in power markets to incentivize adaptation of demand to supply conditions [34]; and wholesale markets are being deregulated and opened to competition [10]. In principle, deregulation can increase efficiency. In practice, the California energy crisis (2001) has demonstrated the high-costs of failure due to flawed deregulation [30, 4], and the importance of testing new market structures in simulation before deploying them [35]. This is the focus of the Power Trading Agent Competition (Power TAC) [13].

In Power TAC, autonomous broker agents compete with each other to make profits in a realistic, detailed smart-grid simulation environment with wholesale, retail and balancing power markets, and about 57,000 customers. The stability of electricity grids critically depends on having balanced electricity supply and demand at all times. Broker agents are financially incentivized to maintain supply-demand balance in their portfolio and thus contribute to grid stability.

It is likely that autonomous broker agents will be employed in future power markets, due to the need to act continually and make real-time decisions in a complex, competitive, dynamic environment. The decision-making challenges of such brokers have been under study in the autonomous agents community, but under either limited scope, or limited competitiveness and comparability [13]. This paper focuses on the question: how should an autonomous broker agent act in competitive power markets to maximize its profits? we advance the state of the art towards answering this question in the following ways:

- This paper is the first to formalize the *complete* broker's power trading problem. We formalize the problem as a Markov Decision Process (MDP) which, due to its continuous high-dimensional state and action spaces, cannot be solved exactly in practice. Our formalization compactly captures the challenges faced by a broker, and provides a guideline for approximating the solution and for extending existing solutions. While our formalization is based on the Power TAC simulation, we expect it to generalize and be useful in reality, since Power TAC closely models real-world markets.

- We present TacTex'15, which is by many metrics the best Power TAC broker at the current time. Using three strategic improvements, TacTex'15 extends a previous strategy which can be viewed as a lookahead-policy [23] that approximates the solution to the MDP. The strategic improvements may seem minor on the surface but result in large performance improvements.

- Using thousands of experiments, we analyze the performance of TacTex'15, and the reasons for its success. Importantly, we investigate how the accuracy of the transition-function predictors (i.e. the demand and cost predictors) affect the performance of the broker's power trading lookahead-policy.

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*

This paper's contributions can be valuable for two communities. For the artificial intelligence community, this paper is a case-study of effective sequential decision making in a domain too complex to be solved by current out-of-the-box methods, which involves autonomous agents, multi-agent systems, game-theoretic considerations, markets, and a potentially important computational sustainability application [7]. For the power markets community, this paper may provide insights on autonomous trading in modern electricity markets, due to the realism of the Power TAC simulation environment.

2. POWER TAC GAME DESCRIPTION

Power TAC is an annual competition in which the competitors are autonomous brokers programmed by teams from around the world. The competition includes hundreds of games and takes several days to complete. In a game, the Power TAC simulator runs on a central server, while competing brokers run remotely and communicate with the server through the internet. Each broker receives partial state information from the server, and responds by communicating the actions it takes. The competition includes different game sizes, ranging from small to large number of competitors. Participants release their broker binaries after the competition, and use them to run controlled experiments.

Power TAC uses a high-fidelity power markets simulator, modeling a smart-grid with more than 57,000 simulated customers (50,000 consumers and 7,000 renewable producers). Power TAC's customers are *autonomous agents* that optimize the electricity-costs and comfort of their human owners [28]. Customers model commercial and residential buildings, solar/wind farms, storage facilities and electric vehicles. Customers consume/produce using time-series generators constructed from real-world data, according to weather and calendar factors. The simulation proceeds in 1-hour timeslots for 60 simulated days and completes in 2 hours.

Figure 1 shows the structure of the Power TAC simulation environment. In Power TAC, autonomous broker agents compete by acting in three markets: (1) a *wholesale market*, where brokers bid in sequences of 24 double auctions to procure energy from generation companies (or sell surplus), to be delivered in the following 24 hours, (2) a *tariff market*, which is a retail market where energy is traded with consumers and distributed renewable energy producers, and (3) a *balancing market*, which ensures that supply and demand are balanced and determines the broker imbalance fees.

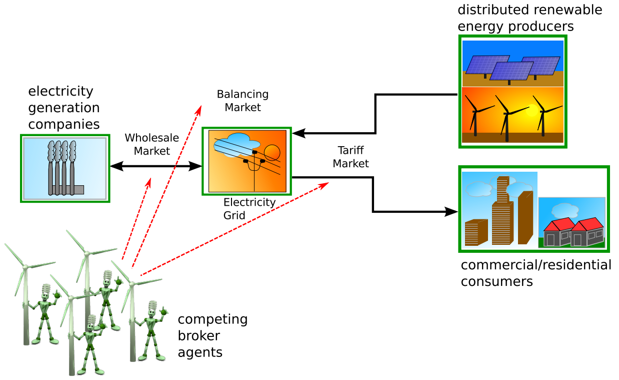

Figure 1: High-level structure of the Power TAC simulation

The brokers compete to gain market share and maximize profit by trading electricity. In the tariff market, brokers publish tariff contracts for energy consumption/production. Tariffs may include fixed and varying prices and possibly bonuses/fees. Customers stochastically subscribe to tariffs which maximize their utility (cost savings and comfort). Customers are equipped with smart-meters, so consumption and production are reported to the broker every hour. Brokers typically balance their portfolio's net demand by buying in the wholesale market. Full details can be found in the Power TAC Game Specification [12].

3. THE BROKER'S POWER TRADING PROBLEM

This section formalizes the broker's power trading problem. Our formalization compactly captures the complex challenges faced by a broker, and provides a guideline for approximating the solution and for extending existing solutions. While our formalization is based on the Power TAC simulator, we expect it to generalize and be useful in reality, since Power TAC closely models real-world markets. We start with an intuitive problem description and continue to our formalization.

Figure 2 illustrates the temporal structure of a broker's power trading problem. The temporal structure of the tariff and wholesale market actions differ in multiple ways. Tariffs specify energy for *immediate* and *repeated* delivery and are published at *low-frequency* (every one or more days). Wholesale bids typically specify energy for *future*, *one-time* delivery and are executed at *high-frequency* (every hour).

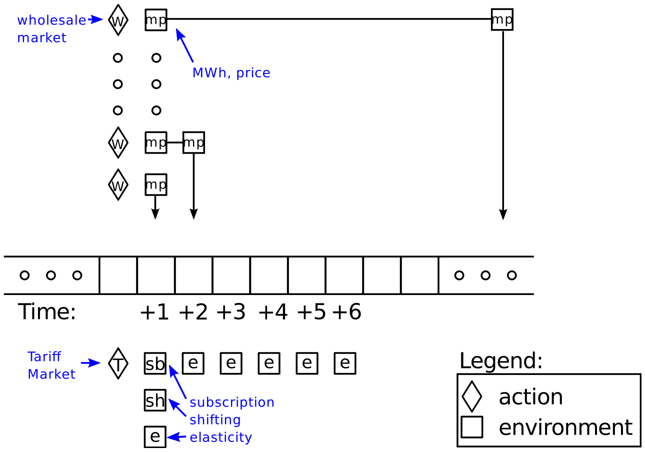

Figure 2: **Temporal structure of the power trading problem.** Time progresses to the right; the notation '+i' stands for 'i timeslots into the future'. Diamonds stand for broker actions. Squares stand for simulation environment responses. The top part represents the wholesale market: a broker submits limit orders to buy/sell energy for the next 24 hours, then it receives the results of the 24 double-auctions. The bottom part represents the tariff market: a broker may publish one or more tariffs (once every 6 hours), and customers respond by potentially (1) subscribing to new tariffs, (2) shifting consumption to cheaper times, and (3) elastically adapting total consumption based on price.

Power Trading as an MDP.

Given the internal states of the simulator and competing brokers, the broker's energy trading problem is a Markov Decision Process (MDP) [25]. However, since competitors' states and parts of the simulator state are unobservable, the trading problem is a Partially Observable MDP (POMDP). Nevertheless, for computational tractability and modeling clarity, we treat unobservable parts of the state as environment stochasticity and formulate the trading problem as an MDP, as follows (denoting B_0 as the acting broker):

- **States:** S is the set of states, where state s is a tuple $\langle t, \mathcal{B}, \mathcal{C}, \mathcal{P}, \mathcal{T}, \mathcal{S}_{B_0}, \mathcal{Q}_{B_0}, \mathcal{A}_{B_0}, I_{B_0}, \mathcal{W}, cash_{B_0}, \rho\rangle$ that includes the current time t (which encapsulates weekday/hour), and the sets: competing broker identities $\mathcal{B}$; identities of consumers $\mathcal{C}$ and producers $\mathcal{P}$ (both referred to as *customers*); published tariffs of all brokers $\mathcal{T} := \cup_{B\in\mathcal{B}}\mathcal{T}_B$; customer subscriptions to B_0's tariffs $\mathcal{S}_{B_0}$; current energy consumption/production of B_0's customers $\mathcal{Q}_{B_0}$; recent auction results $\mathcal{A}_{B_0} := \{\langle p^c, q^c, {\mathcal{O}^c}_{B_0}, \mathcal{O}^u, \mathcal{M}_{B_0}\rangle_j\}_{j=t+1}^{t+24}$ including, for each of the following 24 timeslots, the clearing price p^c and total quantity q^c, B_0's cleared orders ${\mathcal{O}^c}_{B_0}$, all brokers' uncleared orders $\mathcal{O}^u$, and B_0's *market-positions* $\mathcal{M}_{B_0}$ (energy deliveries and charges, updated incrementally from ${\mathcal{O}^c}_{B_0}$); B_0's energy imbalance I_{B_0}; current weather and forecast $\mathcal{W}$; B_0's cash balance $cash_{B_0}$; and randomly sampled game-parameters (such as fees and game length) ρ. **Note:** the underlying state of the game, which includes elements unobserved by the broker, is the tuple $\langle t, \mathcal{B}^o, \mathcal{G}^o, \mathcal{C}^o, \mathcal{P}^o, \mathcal{T}, \mathcal{S}, \mathcal{Q}, \mathcal{A}, \mathcal{I}, \mathcal{W}, cash, \rho\rangle$ where $\mathcal{B}^o, \mathcal{G}^o, \mathcal{C}^o, \mathcal{P}^o$ are the identities *and* internal states of brokers, generation companies, consumers and producers, respectively; and where $\mathcal{S} := \cup_{B\in\mathcal{B}}\mathcal{S}_B$, $\mathcal{Q} := \cup_{B\in\mathcal{B}}\mathcal{Q}_B$, $\mathcal{A} := \cup_{B\in\mathcal{B}}\mathcal{A}_B$, $\mathcal{I} := \{I_B\}_{B\in\mathcal{B}}$, $cash := \{cash_B\}_{B\in\mathcal{B}}$.

- **Actions:** A broker's set of actions $A := A^\tau \cup A^\omega \cup A^\beta$ is composed of tariff market actions A^τ, wholesale market actions A^ω and balancing market actions A^β, as follows.

  1. *Tariff market actions* A^τ: create/modify/revoke tariffs. A tariff is a tuple $T = \langle type, rates, fees\rangle$ where:
     - *type* $\in$ {consumption, production,...} can be general (e.g. production) or specific (e.g. solar-production).
     - *rates*: a set of rates, each specifying a price and when it applies (times and/or usage thresholds).
     - *fees*: optional periodic/signup/withdraw payments.
  2. *Wholesale market actions* A^ω: submit limit orders of the form

     $$\langle energyAmount, limitPrice, targetTime\rangle$$

     to buy/sell energy for one of the next 24 hours.
  3. *Balancing market actions* A^β: submit customers energy curtailment requests (currently unused).

- **Transition Function:** The transition function is partially deterministic and partially stochastic, as follows. The time t is incremented by 1 hour; $\mathcal{B}, \mathcal{C}, \mathcal{P}$ remain unchanged; $\mathcal{T}$ is updated by create/modify/revoke tariff actions, deterministically by B_0, and stochastically (due to unobservability) by other brokers; $\mathcal{S}_{B_0}$ is updated stochastically based on customers' decisions; $\mathcal{Q}_{B_0}$ is determined stochastically based on weather and customers' internal states (shifting and elasticity, see Figure 2); $\mathcal{A}_{B_0}$ is updated with auction results, stochastically since (i) competitors rely on stochastic information (demand predictions), (ii) competitors' internal states are hidden, and (iii) generation companies bid stochastically; I_{B_0} is a deterministic function of $\mathcal{T}_{B_0}, \mathcal{S}_{B_0}, \mathcal{Q}_{B_0}, \mathcal{A}_{B_0}$; $\mathcal{W}$ is stochastic; *cash* is updated deterministically from the recent stochastic reward; and ρ remains unchanged.

- **Reward:** Let s_t, r_t, a_t be the state, reward, and broker-action(s) at time t. Let $r^\tau, r^\omega, r^\beta$ be the broker's energy buy/sell payments in the tariff, wholesale, and balancing markets respectively. Let *dist* be the energy distribution fees, and *fees* the tariff-market fees. The reward at time t can be characterized by the following function.

  $$\begin{aligned} r_t(s_{t-1}, a_{t-1}, s_t) &:= r^\tau(s_t) + r^\omega(s_t) + r^\beta(s_t) \\ &+ dist(s_t) + fees(s_{t-1}, a_{t-1}, s_t) := \\ &\underbrace{Q_t^{cons}p_t^{cons} - Q_t^{prod}p_t^{prod}}_{r^\tau(s_t)} + \underbrace{Q_t^{ask}p_t^{ask} - Q_t^{bid}p_t^{bid}}_{r^\omega(s_t)} \\ &\underbrace{\pm bal(I_{B_0,t})}_{r^\beta(s_t)} \underbrace{-max(Q_t^{cons}, Q_t^{prod}) \times distFee}_{dist(s_t)} \\ &\underbrace{-pub(a_{t-1}) - rev(a_{t-1}) \pm psw(\mathcal{S}_{B_0,t-1}, \mathcal{S}_{B_0,t})}_{fees(s_{t-1},a_{t-1},s_t)} \end{aligned} \quad (1)$$

  where $\pm$ denotes components that can be positive or negative; Q_t^{cons}, Q_t^{prod} are the total consumed/produced quantities by B_0's customers in the tariff-market (both are sums of entries of $\mathcal{Q}_{B_0}$); Q_t^{ask}, Q_t^{bid} are the amounts B_0 sold/procured in the wholesale-market (both are sums of elements of $\mathcal{M}_{B_0}$ inside $\mathcal{A}_{B_0}$); p_t^{cons}, p_t^{prod}, p_t^{ask}, p_t^{bid} are the average buying/selling prices (determined by $\mathcal{T}_{B_0}$, $\mathcal{S}_{B_0}$, $\mathcal{Q}_{B_0}$ and $\mathcal{M}_{B_0}$); $bal(I_{B_0,t})$ is the fee for imbalance $I_{B_0,t} = Q_t^{cons} - Q_t^{prod} + Q_t^{ask} - Q_t^{bid}$ (which depends on unobserved other broker imbalances $I \backslash I_{B_0,t}$); $distFee$ is a fixed fee per kWh transferred over the grid; pub, rev are tariff publication and revoke fees; psw are tariff periodic/signup/withdraw fees/bonuses.

- **Discount Factor:** γ reflects daily interest on cash balance.

4. THE TacTex'15 BROKER AGENT

This section characterizes approximate MDP solutions, and describes TacTex'15's approximate solution.

4.1 Approximate MDP Solutions

The MDP's solution is an optimal power-trading policy (a mapping from states to actions). There are two problems to solve the MDP exactly: first, the high-dimensional states and actions and the complex reward makes it computationally intractable, and second, some components of the transition and reward functions are unknown to the broker. Therefore, brokers necessarily can only approximate the solution. There are four categories of approximate solutions to large MDPs, of which *lookahead policies* seem suitable for our domain, since they are effective in time-varying settings, where it is unclear how to approximate a value function or find a simple rule that maps states to actions [23].

Lookahead policies are partial MDP solutions that have been effective in high-dimensional state-spaces [24, 3, 11, 29, 5, 6, 18, 32]. Lookahead policies optimize over simulated trajectories s_t, r_t, a_t, s_{t+1}, r_{t+1}, a_{t+1},... using generative models that predict *action effects* (next state and reward). Here, the reward is a deterministic function of s_{t-1}, a_{t-1}, s_t except for the $bal(I_{B_0,t})$ component. Therefore a broker needs generative models for $bal(I_{B_0,t})$, for $\mathcal{T} \setminus \mathcal{T}_{B_0}, \mathcal{S}_{B_0}, \mathcal{Q}_{B_0}$ (to predict Q_t^{cons}, p_t^{cons}, Q_t^{prod}, p_t^{prod}), and for $\mathcal{A}_{B_0}$ (to predict Q_t^{ask}, p_t^{ask}, Q_t^{bid}, p_t^{bid}).

While these action effects can be predicted independently, actions need to be optimized in conjunction: the $bal(I_{B_0,t})$ function is designed such that imbalance fees typically result in negative reward when taking actions of a single type, while positive reward can be achieved by taking actions of multiple types in parallel (to maintain low imbalance). Therefore, any tractable lookahead policy is required to efficiently (i) sample, and (ii) combine the actions to simulate.

The 2013 champion, TacTex'13 [33], can be viewed as approximating an MDP solution using a lookahead policy. TacTex'13 does not optimize production tariffs, wholesale selling and fees, so $Q_t^{prod}p_t^{prod}$, $Q_t^{ask}p_t^{ask}$, and $psw()$ are always zero in Equation 1. TacTex'13's main routine is roughly Algorithm 1. For each tariff in a sample of fixed-rate consumption tariffs (line 1), it uses a *demand-predictor* to predict $Q_t^{cons}p_t^{cons}$ for each t in the horizon (line 2), assumes $Q_t^{bid} = Q_t^{cons}$ (and therefore marks both as Q_t), uses a *cost-predictor* to predict for every t the price p_t^{bid} of buying Q_t (line 4), predicts a profit (called *utility* or *return*) as the sum of rewards along the horizon (line 5), and executes the utility-maximizing combination of actions (lines 7-8). Therefore, TacTex'13 lookahead efficiently *combines* actions (addressing (ii) from above) by constraining $Q_t^{bid} = Q_t^{cons}$, instead of examining combinations (therefore $bal(I_{B_0,t}) = 0$). TacTex'13 efficiently *samples* actions (addressing (i) from above) by sampling fixed-rate tariffs in a limited region in the tariff market, and by treating wholesale actions hierarchically: it (a) treats Q_t^{bid} as an action to be sampled (in Q_t^{cons} values), and (b) solves a subproblem of finding a cost-minimizing sequential bidding policy $\pi(Q)$ for procuring quantities Q on a small MDP isolated from the full MDP.

Algorithm 1 TacTex'13's Lookahead Policy

```
1: for trf in sampleCandidateTariffs() ∪ {no-op} do
2:     $\{\langle Q_t, p_t^{cons}\rangle | t = +1, ..., +T\} \leftarrow$ demandPredictor.predict(trf)
3:     for t in $\{+1, ..., +T\}$ do
4:         $p_t^{bid} \leftarrow$ costPredictor.predict($Q_t$)
5:     utilities[trf] $\leftarrow \sum_{t=+1}^{+T} Q_t p_t^{cons} - Q_t p_t^{bid} - dist(Q_t) - pub(\text{trf})$
6: trf* $\leftarrow \arg\max_{\text{trf}}$ utilities[trf]
7: publishTariff(trf*) // tariff market action, possibly no-op
8: procure $\{Q_t\}_{t=+1}^{+T}$ predicted for trf* in line 2 // wholesale market
```

4.2 TacTex'15's Architecture

TacTex'15's architecture is similar to that of TacTex'13 in four main ways. TacTex'15 does not try to benefit from (1) production tariffs (2) wholesale selling, (3) imbalance, and (4) tariff fees, since in preliminary tests (1)-(3) did not seem beneficial and (4) had some simulator implementation issues (see Section 5.1). As a result, TacTex'15 assumes $Q_t^{prod}p_t^{prod}$, $Q_t^{ask}p_t^{ask}$, $bal()$, and $psw()$ in Equation 1 to be zero. Therefore TacTex'15's lookahead policy is quite similar to TacTex'13's (Algorithm 1); we refer the reader to [33] for pseudo-code of the main routines. On the other hand, TacTex'15 introduces three main improvements over TacTex'13: it uses different (1) demand-predictor, (2) cost-predictor (both (1)-(2) are *transition-function* predictors), and (3) wholesale bidding strategy π.

Demand Predictor. The demand-predictor predicts customer subscription changes and future demand, which determine $Q_t^{cons}p_t^{cons}$. TacTex'13 learned a demand-predictor from data. However, in Power TAC there is no need to do so: these complex stochastic customer behaviors are coded in Power TAC's open-source simulator. Instead, TacTex'15 uses the simulator's customer code as a basis for its demand-predictor. However, this code does not provide a complete demand-predictor: it relies on information hidden from brokers. TacTex'15 seeds this information with expected values: customers of other brokers are assumed to be subscribed to the best tariffs, customer subscriptions are predicted in the limit (expected values after infinite time), and customer demand parameters are set to expected values.

The simulator's customer code is a high-quality demand-predictor which, beyond contributing to the broker's performance, allows us to empirically analyze the dependency of the broker's performance on the demand-predictor's accuracy, and more generally to gain insight into the dependency of lookahead-based power trading on the transition-function predictors' accuracy. Section 5 includes the results of such an analysis which, due to the realism of the Power TAC simulator, could generalize to real markets.

Cost Predictor. Wholesale costs are determined by procured quantities and brokers' bidding strategies, which may change dynamically. TacTex'15 uses an adaptive cost-predictor $Q_t^{bid} \mapsto p_t^{bid}$, described in Algorithm 2. It has two components: a linear regression predictor trained on *boot data* (wholesale transactions sent by the simulator at game start) (line 1), and a real-time correction factor constructed from the last 24 hours' prediction errors (line 2). Since the correction factor is constructed from little data (to ensure responsiveness), we limit it to bias correction. The boot data is larger (336 instances) so we use it to determine the slope. TacTex'13's cost-predictor ignored Q_t^{bid}, and predicted past average prices based on time. We compare the two predictors in Section 5.

Algorithm 2 cost-predictor(Q_t^{bid})

```
1: reg $\leftarrow$ trainLinearRegression($\{\langle Q_i^{bid}, p_i^{bid}\rangle\}_{i \in bootdata}$)
2: correctionFactor $\leftarrow$ averagePredictionErrorInLast24Hours()
3: return reg.predict($Q_t^{bid}$) - correctionFactor
```

Wholesale Bidding Strategy. TacTex'15 uses a combination of truthful and strategic (i.e. non-truthful) bidding. A truthful bidder sets its limit price to the predicted imbalance fee $\overline{p}$. It gets the highest priority among competitors who bid less than $\overline{p}$ and never pays more than $\overline{p}$. However, since the sequential double-auction mechanism is not incentive compatible, truthful bidding is suboptimal in some situations. TacTex'13 used a learned optimistic strategic (i.e. non-truthful) bidding strategy $\pi(Q)$ that assumed that bids with limit-prices higher than a double-auction's clearing price would get fully cleared. This strategy is optimal in some situations (e.g. single-buyer or cooperative setups), but can be exploited by competitors who learn to bid slightly higher. Since each of the two strategies is beneficial in different situations, combining them provides a form of hedging for TacTex'15. Let $\underline{p}$ be the limit price suggested by TacTex'13's strategy, and ϵ be the minimum amount that can be traded (0.01 mWh). To bid for a quantity Q_t^{bid}, TacTex'15 submits the following 25 orders (see MDP wholesale actions, Section 3) $\langle Q_t^{bid} - 24\epsilon, \overline{p}, t\rangle$, $\{\langle \epsilon, \underline{p} + i\frac{\overline{p} - \underline{p}}{24}, t\rangle\}_{i=0}^{23}$. This strategy benefits from both worlds: if TacTex'15 sets the price, it will either be the strategic price returned by $\pi(Q)$, or the lowest among its higher bids. If another broker sets the price, TacTex'15 will have a higher priority and benefit from the lower price as long as it is not higher than $\overline{p}$.

Future Extensions of TacTex'15's MDP Solution. Referring back to the reward specification (Equation 1), our MDP provides a guideline for future extensions of TacTex'15's lookahead policy. In some situations a broker can profit from imbalance. We can relax the assumption that $Q_t^{cons} = Q_t^{bid}$, add imbalanced trajectories to our lookahead search, setting $Q_t^{cons} - Q_t^{bid} = I_{B_0,t}$ for a sample of $I_{B_0,t}$ values, and predict $bal(I_{B_0,t})$ using a learned predictor. We can sample production tariffs like consumption tariffs, and treat wholesale sell hierarchically like wholesale buy actions. This addresses requirement (i) from Section 4.1 (sample actions efficiently). However, addressing requirement (ii) (combined actions efficiently) becomes more challenging. In an initial implementation we use an alternating, local improvement based approach which performs well, but more sophisticated methods might be possible. Finally, tariff-revoke actions can be added by simulating lookahead trajectories with each of the active tariffs removed. Initial implementation shows promising results.

5. RESULTS

We analyze TacTex'15's performance in competitions (Section 5.1) and controlled experiments (Section 5.2).

5.1 Competition Results

The Power TAC 2015 Finals included 11 teams from universities in America, Europe and Asia. 230 games were played continually over a week, in three different sizes: 3-brokers, 9-brokers, and 11-brokers. A day after the finals ended, 8 of the teams competed in a post-finals, demo-competition with 70 4-broker games. While being unofficial, this competition was run similarly to the finals with one important difference: a simulator-loophole that was exploited during the finals, was fixed. Due to the proximity to the finals, and a parallel workshop, we believe that teams used the same brokers they used in the finals.

Table 1 summarizes the 2015 finals results. While TacTex'15 was officially ranked 2nd, it was the best broker that did not exploit a simulator-loophole: the 1st-ranked broker gained the highest overall score by exploiting a simulator loophole in 3-broker games, which resulted in unrealistic dynamics and an unrealistically high score that biased the final ranking (see dark gray cells in Table 1).[1] Specifically, Maxon15 subscribed customers to inflated tariffs which promised customers large payments if customers *unsubscribed* from them after a period shorter than a single timeslot. While customers had no way to unsubscribe quickly enough to collect these payments, due to the loophole they subscribed to these tariffs assuming they *could* collect the payments, thus paying inflated prices to Maxon15.

Table 1: **Power TAC 2015 finals results.** Ranking is determined by the "Total" score, which is a sum of individual z-scores in each game size, displayed in the columns "11-brokers" (10 games played by all brokers), "9-brokers" (45 games played by each broker) and "3-brokers" (45 games played by each broker).

Broker	11-brokers	9-brokers	3-brokers	Total
Maxon15	0.611	0.801	1.990	3.402
TacTex'15	*0.897*	*1.066*	*0.258*	*2.221*
CUHKTac	0.962	0.859	0.106	1.927
AgentUDE	0.421	0.367	0.809	1.597
Sharpy	0.429	0.614	0.521	1.564
COLDPower	0.726	0.397	-0.751	0.371
cwiBroker	-0.002	-0.120	0.465	0.343
Mertacor	0.413	0.142	-1.341	-0.786
NTUTacAgent	-1.017	-1.638	0.453	-2.202
SPOT	-1.052	-0.243	-1.032	-2.327
CrocodileAgent	-2.387	-2.244	-1.479	-6.111

After the finals, the loophole was fixed. When replaying 3-broker competition games without the loophole, Maxon15 no longer won by a large gap, but instead lost by a large gap to TacTex'15. When taking into account only 11- and 9-broker games from the finals (where the loophole had no impact), TacTex'15 ended 1st with a total z-score of 0.142 ahead of CUHKTac and 0.551 ahead of Maxon15, finishing slightly behind CUHKTac in 11-broker games (by 0.065) and ahead of CUHKTac in 9-broker games (by 0.207). In the post-finals demo competition with a repaired simulator, TacTex'15 won by a large gap ahead of the others (Table 2), making 50% more profits than the 2nd place (Maxon15). Maxon15 used the same strategy as before, but it was not as effective with the loophole fixed.[2]

Table 2: **Power TAC 2015 post-finals demo competition results.** 70 games were played in a single game-size (4-brokers). Ranking is determined by z-score.

Broker	4-brokers (profits)	4-brokers (z-score)
TacTex'15	*15.0M*	*1.122*
Maxon15	10.7M	0.627
CUHKTac	10.0M	0.537
AgentUDE	9.7M	0.509
cwiBroker2015	7.9M	0.297
Sharpy	4.6M	-0.092
COLDPower	-0.8M	-0.724
SPOT	-14.0M	-2.276

Figure 3 shows an analysis of TacTex'15's performance

[1] The loophole's exploitation was confirmed by the competition organizers. However, Maxon was not disqualified: they explained it as an unintended result of automatic parameter tuning right before the finals.

[2] To be fair, one should note that they did not retune their parameters to the repaired simulator. On the other hand, it's not clear that other parameters would have done particularly better in the absence of the loophole.

in the 2015 finals and in the post-finals competition. In 11-broker games CUHKTac (1st) and TacTex'15 (2nd) won by a large gap over the other brokers, where most brokers ended with losses. In 9-broker games TacTex'15 won by a large gap, making 30% more profit than the 2nd place broker in this category (CUHKTac), despite missing 3 out of 45 games due to network connection problems. The revenue and costs plots show that in 11- and 9-broker games TacTex'15 chose to reduce its market share, likely due to the fierce competition, so that its revenue and costs were lower compared with other top brokers, while its profit remained high. In 3-broker games TacTex'15 typically performed the best, although this is harder to see in the figure, due to several events that biased the final averages: (a) Maxon15's loophole-exploitation, discussed above; (b) About 1/2 of of AgentUDE's, Sharpy's and cwiBroker's 3-broker game scores come from single outlier games in which they played against a non-functioning broker and/or a competitor's crash in a monopoly/duopoly situation; (c) TacTex'15 missed 5 out of its 45 3-broker games due to network connection problems, resulting in a score of 0 in these games, and a reduction of 4.3% in TacTex'15's average profit. In the 4-broker games of the post-finals competition TacTex'15 made about 50% more profit than the 2nd place broker. The revenue and costs plots show that it had a similar revenue to the 2nd and 3rd place brokers, but much lower costs; higher revenue and lower costs than the 5th, 6th brokers; and almost double the revenue of each of the other brokers.

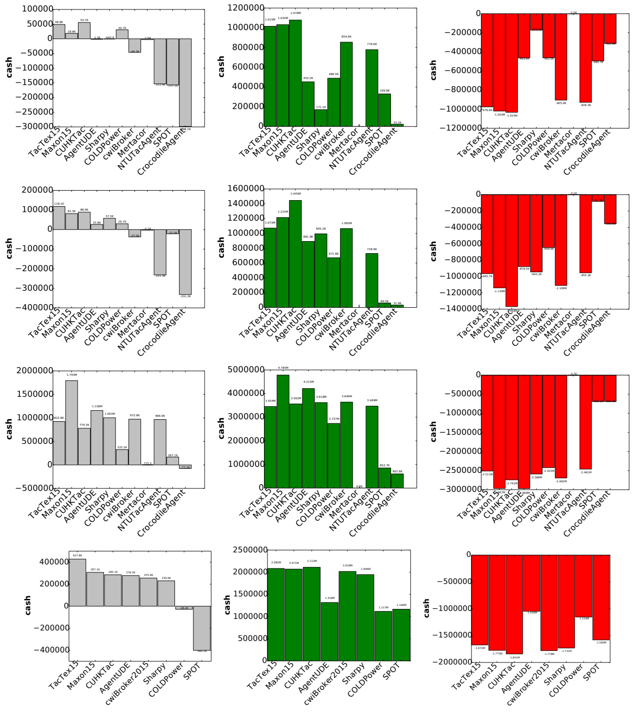

Figure 3: **Competition analysis: average profit, revenue and costs.** The top 3 lines respectively summarize 11-, 9-, 3-broker games from Power TAC 2015 finals; bottom line summarizes the 4-broker games of the post-finals demo competition. Each line shows average profit (left), revenue (middle), and costs (right). 3-broker game results are biased due to a simulator-loophole exploitation by Maxon15 and a few other events (see text for details)

5.2 Controlled Experiments

While the competition is motivating and its results are illustrative, it cannot isolate specific broker components in a statistically significant way. We therefore subsequently tested TacTex'15 in thousands of games, in two types of controlled experiments: (a) performance tests, and (b) ablation analysis tests, which evaluate the contribution of TacTex'15's main components to its overall performance.[3]

5.2.1 *Experimental Setup*

Each experiment consisted of running 56 games against a set of opponent brokers, using broker binaries of 2015 finalists. To better evaluate statistical significance, we held most of the random factors in the simulation fixed across experiments (random seeds, weather conditions). To fix weather conditions, we used weather files containing 3 months of real-world weather. To cover year-round conditions we used 8 weather files (each file used by 1/8 of the games) with start-dates of January, April, July, October of 2009 and 2010.

5.2.2 *Performance Tests*

A successful broker should perform well in expectation against every set of opponents, under different stochastic conditions (here weather/random seeds). At the time of writing this paper, five 2015 finalists have released their brokers' binaries. We used these binaries to test TacTex'15's performance in $2, 3, \ldots, 6$-broker games. We generated combinations of brokers for each game size, and tested each combination in 56 games, as described above. Figure 4 presents the results. TacTex'15 significantly won against every combination of opponents, typically by a large gap.

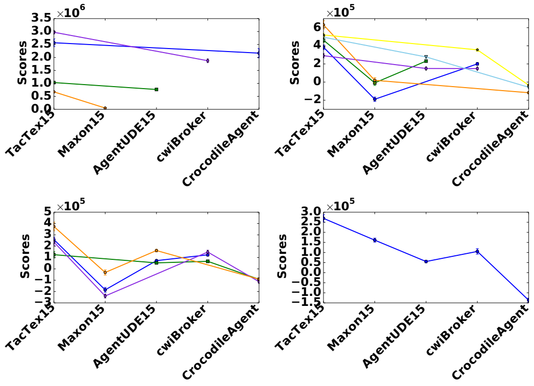

Figure 4: **Performance of TacTex'15 against Power TAC 2015 finalists in controlled experiments of game-sizes of 2-5.** Each line connects n points where each point is an average score (y-value) of a broker (x-value). Therefore, each line represents the average scores of a combination of brokers playing each other under a variety of conditions (note the small error bars). Results are shown for game-sizes of 2-, 3-, 4-, 5-brokers (top-left, top-right, bottom-left, bottom-right, respectively). Similar results for 6-brokers are omitted. TacTex'15 consistently won against all combinations of brokers, in all game-sizes.

[3] To clarify, in all controlled experiments TacTex'15 was the exact competition agent. Its demand predictor has never used the simulator code with the loophole since this code relied on information hidden from the broker.

5.2.3 *Ablation Analysis*

To understand the reasons for TacTex'15's success, we tested the contribution of TacTex'15's main components to its overall performance, in all possible game-sizes (2,...,6). We created three ablated versions of TacTex'15 by disabling each of its main components. For each game size, we selected the "strongest" combination of opponents, against which TacTex'15 had the lowest score. We tested each ablated version against these opponents in a 56-game experiment, holding random seeds and weather conditions fixed to the same values used against TacTex'15. When disabling a component, we used as a baseline the corresponding component used by TacTex'13 (since TacTex'15's ablated version must have some component in place of a disabled one to run properly). Figure 5 shows the results of our ablation analysis. Disabling the cost-predictor (Abl-cost) did not have significant impact on TacTex'15's performance (however it can reduce performance, see Figure 7). Disabling the wholesale-bidding strategy (Abl-bid) significantly hurts TacTex'15's performance: it reduces TacTex'15's score in game sizes 2, 4, 5, 6, and it causes TacTex'15 to either lose its lead (in game sizes 2, 3) or have a smaller victory margin (in game sizes 4, 5, 6). Disabling the demand-predictor (Abl-demand) significantly hurts TacTex'15's performance: it drops TacTex'15's score in all game sizes, and causes TacTex'15 to either lose its lead (in game sizes 3, 5, 6) or have a smaller victory margin (in game sizes 2, 4).

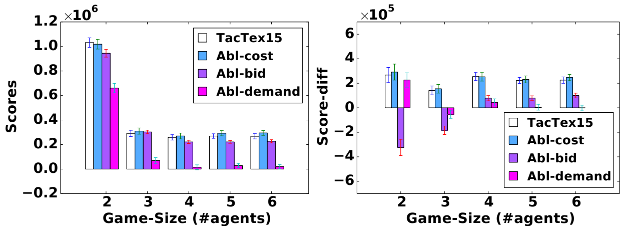

Figure 5: **Ablation analysis for 2-6 broker games.** The performance of TacTex'15 is compared with three of its ablated versions, when playing against the strongest combination of opponents in each game size. Ablated versions are constructed from TacTex'15 by disabling cost predictor (Abl-cost), wholesale-bidding strategy (Abl-bid), and demand-predictor (Abl-demand). The left figure shows the average scores of each version in each game size; the right figure shows the average score-differences of each version from opponents' average score (y-axes' scales are the same).

5.2.4 *Ablation Analysis Extensions*

To gain more insight into the importance of TacTex'15's main components, we extended each ablation experiment. First, we extended TacTex'15's demand-predictor ablation analysis from a binary ablation test (disabled/enabled, see Abl-demand in Figure 5) to a continuum of ablation-levels, thus testing TacTex'15's sensitivity to demand prediction errors. Figure 6 shows the performance-degradation as a function of ablation-level. We see that TacTex'15's degrades quickly even for small levels of ablation. We conclude that having an accurate demand-predictor is crucial for TacTex'15's success.

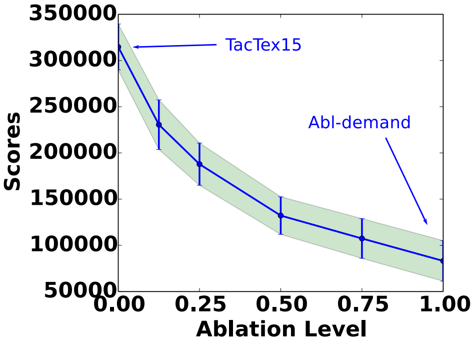

Figure 6: **Performance as a function of ablation level of the demand-predictor in 3-agent games.** The plot shows the degradation in TacTex'15's performance as the ablation level of its demand predictor increases. To change ablation level along a continuum, TacTex'15 uses here a weighted combination of two demand-predictors: (1) its own predictor, and (2) TacTex'13's demand-predictor, which was used by the ablated agent Abl-demand in Figure 5. Ablation level is then represented as the relative weight given to predictor (2), so that a weight of 0 means "no-ablation", and a weight of 1 means "full-ablation".

Next, we extended the ablation analysis of TacTex'15's wholesale-bidding strategy with additional comparisons against its ablated version (used by Abl-bid, see Figure 5). Abl-bid's strategy (which is TacTex'15's strategy) can be viewed as more cooperative than TacTex'15's, since it submits lower bids, and thus may result in lower costs against an opponent using a similar strategy. To understand whether Abl-bid's cooperative strategy is preferable in some situations, we created a payoff matrix (Table 3) by running 2-broker games, testing both TacTex'15 and Abl-bid in self-play and against each other. While Abl-bid's cooperative strategy indeed resulted in lower costs in self-play (40 \$/mWh vs. 57 \$/mWh, a 29.8% reduction), Abl-bid's total scores in self-play were not higher than TacTex'15's, since the competitive selling strategy reduced selling-prices further than TacTex'15's, such that the profit remained similar to TacTex'15's. As a result, TacTex'15's competitive strategy dominated Abl-bid's cooperative strategy in Table 3's experiments.

Table 3: **Payoff matrix of two wholesale-bidding strategies in 2-agent games.** The matrix shows a game-theoretic payoff matrix of two wholesale bidding strategies: (a) Comp-Bid is TacTex'15's competitive bidding strategy, and (b) Coop-Bid is Abl-bid's (and TacTex'13's) cooperative bidding strategy from Figure 5. The matrix entries show the average scores of agents using these strategies (TacTex'15 and Abl-bid, respectively) in self-play and against each other.

Payoff Matrix

	Coop-Bid	Comp-Bid
Coop-Bid	$1.0M$ $1.0M$	$1.6M$ $0.8M$
Comp-Bid	$0.8M$ $1.6M$	$1.0M$ $1.0M$

We ran additional self-play experiments using 3-, 4-, 5-broker games. In these cases Abl-bid's more cooperative bidding policy resulted in higher scores than TacTex'15, mainly since Abl-bid's lower energy costs enabled a longer price-reduction period after game-start, during which selling-prices where higher than the eventual equilibrium after which the profit of all brokers increased in the same pace.

Finally, we extended TacTex'15's cost-predictor ablation analysis. Even though ablating TacTex'15's cost predictor did not reduce performance against the 2015 finalists (Figure 5), we expect it to reduce performance when wholesale costs change more dynamically. Figure 7 shows the result of such an experiment, where TacTex'15 played against its cost-predictor ablated version (Abl-cost from Figure 5), and was quicker to react to a drop in wholesale costs and thus significantly won against Abl-cost.

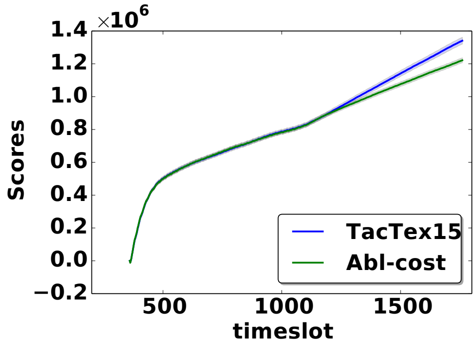

Figure 7: **Cost-predictor ablation in presence of abruptly changing market-costs.** The plot shows the average cumulative profit (with confidence bounds) as a function of time in head-to-head games of TacTex'15 vs. its cost-predictor ablated version (Abl-cost from Figure 5), when market costs abruptly dropped in timeslot 1080 (mid-game). TacTex'15 was quicker to react due to its more adaptive cost-predictor: it reduced selling prices, and thus gained market-share and increased its profits. To create a market-cost drop effect, we could reduce either the sellers' asks, or the brokers' bids. We implemented the latter (to avoid changing the simulator), by making both brokers switch their bidding policies in timeslot 1080 from competitive policies (of TacTex'15) to a cooperative policy (of TacTex'13).

6. RELATED WORK

This work is the first to formalize the *complete* broker's power trading problem as an MDP, and characterize its approximate solutions. Previous research either did not formulate the trading problem explicitly, or used an MDP to model either a more *abstract* trading problem [26, 27], or a *subproblem* of the complete trading problem [22, 16, 33, 15, 1, 20]. Moreover, all these MDP models were *heuristically* and *manually* constructed. In contrast, our MDP is defined by the underlying problem.

Previous approaches to power trading did not use lookahead policies to optimize the predicted utility (other than TacTex'13, which was discussed above). AgentUDE14 [20] (1st place, 2014) used an empirically tuned tariff strategy provoking subscription changes and withdraw payments, and Q-learning for wholesale bidding. CwiBroker14 (2nd place, 2014) [9] used tuned heuristics based on domain knowledge. An analysis of the 2014 Power TAC finals can be found at [2]. Mertacor13 [19] used Particle Swarm Optimization based tariff strategy. CwiBroker13 [17] (2nd place, 2013) used tariff strategy inspired by Tit-for-Tat. Their wholesale strategy used multiple bids per auction but was based on equilibria in continuous auctions, rather than TacTex'15's hedging between optimistic strategic bidding and truthful bidding.

In other trading agent competitions, utility-optimization approaches were used in different market structures [31, 21]. Other approaches included game theoretic analysis of the economy [14] and fuzzy reasoning [8].

7. CONCLUSION

This paper has focused on the question: how should an autonomous electricity broker agent act in competitive electricity markets to maximize its profit. We have formalized the complete electricity trading problem as an MDP, which is computationally intractable to solve exactly. Our formalization provides a guideline for approximating the MDP's solution, and for extending existing approximate solutions. We introduced TacTex'15 which extends the lookahead policy of a previously champion agent (TacTex'13) with better transition function predictors and a more competitive bidding strategy, and achieves state-of-the-art performance in competitions and controlled experiments. Using thousands of experiments against 2015 finalist brokers, we analyzed TacTex'15's performance and reasons for its success, finding that while its lookahead policy is an effective solution in the power trading domain, its performance can be sensitive to errors in the transition function prediction, especially demand-prediction. An important direction for future work is to further close the gap between the current approximate solution to the trading MDP and its fully optimal solution.

Acknowledgments

This work has taken place in the Learning Agents Research Group (LARG) at UT Austin. LARG research is supported in part by NSF (CNS-1330072, CNS-1305287), ONR (21C184-01), and AFRL (FA8750-14-1-0070), and AFOSR (FA9550-14-1-0087).

REFERENCES

[1] 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)*, May 2014.

[2] J. Babic and V. Podobnik. An analysis of power trading agent competition 2014. In *Agent-Mediated Electronic Commerce*, 2014.

[3] D. Bertsekas and D. Castanon. Rollout algorithms for stochastic scheduling problems. *Journal of Heuristics*, 5(1):89–108, 1999.

[4] S. Borenstein. The trouble with electricity markets: Understanding california's restructuring disaster. *Journal of Economic Perspectives*, 16(1):191–211, 2002.

[5] H. Finnsson and Y. Björnsson. Simulation-based approach to general game playing. In *Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 1*, AAAI'08, pages 259–264. AAAI Press, 2008.

[6] S. Gelly and D. Silver. Combining online and offline knowledge in uct. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, pages 273–280, New York, NY, USA, 2007. ACM.

[7] C. P. Gomes. Computational sustainability: Computational methods for a sustainable environment, economy, and society. *The Bridge*, 39(4):5–13, 2009.

[8] M. He, A. Rogers, E. David, and N. R. Jennings. Designing and evaluating an adaptive trading agent for supply chain management applications. In H. L. Poutré, N. Sadeh, and J. Sverker, editors, *Agent-mediated Electronic Commerce, Designing Trading Agents and Mechanisms: AAMAS 2005 Workshop AMEC 2005, Utrecht, Netherlands, July 25, 2005, and IJCAI 2005 Workshop TADA 2005, Edinburgh, UK, August 1, 2005, Selected and Revised Papers*, pages 35–42. Springer, 2005. Event Dates: Auguest 2005.

[9] J. Hoogland and H. L. Poutre. An effective broker for the power tac 2014. In *AAMAS Workshop on Agent-Mediated Electronic Commerce and Trading Agents Design and Analysis (AMEC/TADA 2015)*, May 2015.

[10] P. L. Joskow. Lessons learned from electricity market liberalization. *The Energy Journal*, Volume 29, 2008.

[11] M. J. Kearns, Y. Mansour, and A. Y. Ng. Approximate planning in large pomdps via reusable trajectories. In *Advances in Neural Information Processing Systems 12, [NIPS Conference, Denver, Colorado, USA, November 29 - December 4, 1999]*, pages 1001–1007, 1999.

[12] W. Ketter, J. Collins, P. P. Reddy, and M. D. Weerdt. The 2015 power trading agent competition. *ERIM Report Series Reference No. ERS-2015-001-LIS*, 2015.

[13] W. Ketter, M. Peters, and J. Collins. Autonomous agents in future energy markets: The 2012 Power Trading Agent Competition. In *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*. AAAI, 2013.

[14] C. Kiekintveld, Y. Vorobeychik, and M. Wellman. An analysis of the 2004 supply chain management trading agent competition. In H. PoutrÃl', N. Sadeh, and S. Janson, editors, *Agent-Mediated Electronic Commerce. Designing Trading Agents and Mechanisms*, volume 3937 of *Lecture Notes in Computer Science*, pages 99–112. Springer Berlin Heidelberg, 2006.

[15] 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 *Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2014 IEEE PES*, pages 1–6, Oct 2014.

[16] R. T. Kuate, M. He, M. Chli, and H. H. Wang. An intelligent broker agent for energy trading: An mdp approach. In *The 23rd International Joint Conference on Artificial Intelligence*, 2013.

[17] B. Liefers, J. Hoogland, and H. L. Poutre. A successful broker agent for power tac. In *AAMAS Workshop on Agent-Mediated Electronic Commerce and Trading Agents Design and Analysis (AMEC/TADA 2014)*, May 2014.

[18] R. Lorentz. Amazons discover monte-carlo. In H. van den Herik, X. Xu, Z. Ma, and M. Winands, editors, *Computers and Games*, volume 5131 of *Lecture Notes in Computer Science*, pages 13–24. Springer Berlin Heidelberg, 2008.

[19] E. Ntagka, A. Chrysopoulos, and P. A. Mitkas. Designing tariffs in a competitive energy market using particle swarm optimization techniques. In *AAMAS Workshop on Agent-Mediated Electronic Commerce and Trading Agents Design and Analysis (AMEC/TADA 2014)*, May 2014.

[20] S. Ozdemir and R. Unland. Agentude: The success story of the power tac 2014's champion. In *AAMAS Workshop on Agent-Mediated Electronic Commerce and Trading Agents Design and Analysis (AMEC/TADA 2015)*, May 2015.

[21] D. M. Pardoe. *Adaptive Trading Agent Strategies Using Market Experience*. PhD thesis, 2011.

[22] 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, 2013.

[23] W. Powell and S. Meisel. Tutorial on stochastic optimization in energy – part ii: An energy storage illustration. *Power Systems, IEEE Transactions on*, PP(99):1–8, 2015.

[24] W. B. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality, 2nd Edition*. Wiley, 2011.

[25] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1994.

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

[27] 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 Volume Two*, pages 1446–1451. AAAI Press, 2011.

[28] P. P. Reddy and M. M. Veloso. Factored Models for Multiscale Decision Making in Smart Grid Customers. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI-12)*, 2012.

[29] D. Silver and J. Veness. Monte-carlo planning in large pomdps. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 2164–2172. Curran Associates, Inc., 2010.

[30] S. Stoft. *Index*, pages 460–468. Wiley-IEEE Press, 2002.

[31] P. Stone, R. E. Schapire, M. L. Littman, J. A. Csirik, and D. McAllester. Decision-theoretic bidding based on learned density models in simultaneous, interacting auctions. *Journal of Artificial Intelligence Research*, 19:209–242, 2003.

[32] D. Urieli and P. Stone. A learning agent for heat-pump thermostat control. In *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, May 2013.

[33] D. Urieli and P. Stone. Tactex'13: A champion adaptive power trading agent. In *Proceedings of the Twenty-Eighth Conference on Artificial Intelligence (AAAI 2014)*, July 2014.

[34] U.S. Department of Energy. *"Grid 2030" A National Vision For Electricity's Second 100 Years*, 2003.

[35] A. Weidlich and D. Veit. A critical survey of agent-based wholesale electricity market models. *Energy Economics*, 30(4):1728 – 1759, 2008.