In *Proceedings of the 28th Conference on Artificial Intelligence* (AAAI 2014), Quebec City, Quebec, Canada, July 2014.

TacTex'13: A Champion Adaptive Power Trading Agent

Daniel Urieli and **Peter Stone**
Dept. of Computer Science
The University of Texas at Austin
Austin, TX 78712 USA
{urieli,pstone}@cs.utexas.edu

Abstract

Sustainable energy systems of the future will no longer be able to rely on the current paradigm that energy supply follows demand. Many of the renewable energy resources do not produce power on demand, and therefore there is a need for new market structures that motivate sustainable behaviors by participants. The Power Trading Agent Competition (Power TAC) is a new annual competition that focuses on the design and operation of future retail power markets, specifically in smart grid environments with renewable energy production, smart metering, and autonomous agents acting on behalf of customers and retailers. It uses a rich, open-source simulation platform that is based on real-world data and state-of-the-art customer models. Its purpose is to help researchers understand the dynamics of customer and retailer decision-making, as well as the robustness of proposed market designs. This paper introduces TACTEX'13, the champion agent from the inaugural competition in 2013. TACTEX'13 learns and adapts to the environment in which it operates, by heavily relying on reinforcement learning and prediction methods. This paper describes the constituent components of TACTEX'13 and examines its success through analysis of competition results and subsequent controlled experiments.

1 Introduction

Sustainable energy systems of the future will have to include a robust solution to a major challenge presented by many of the renewable energy resources (wind, solar, tidal, etc.): these resources do not produce power on demand. As a result, energy consumption patterns will have to adapt to the availability of renewable energy supply (Ramchurn et al. 2012). This creates a need for new market structures that financially incentivize desired consumer behaviors, such as shifting consumption to times when more energy is available, and utilizing distributed storage and small-scale production technologies more effectively (Ketter, Peters, and Collins 2013). Indeed, governments around the world are acting to re-engineer their electricity grid into a smart-grid with supporting retail market infrastructure and customer participation in power markets through demand-side management and distributed generation (U.S 2003; Eur 2011). As a part of this process, energy markets are being opened to competition, however, the transition to competitive markets can be risky (Borenstein 2002).

The Power Trading Agent Competition (Power TAC) is a low-risk platform for modeling and testing retail power market designs and related automation technologies. It simulates a future smart grid environment with renewable energy production, smart metering, autonomous agents acting on behalf of customers and retailers, state-of-the-art customer models, and realistic market designs. Since wholesale markets are not designed for individual customer participation, retail brokers can serve as financial intermediaries, representing large number of customers and thus minimizing risk-adjusted costs, so that they make profit while reducing energy prices for their customers (Ketter, Collins, and Reddy 2013). In Power TAC, several self-interested, autonomous broker agents compete with each other with the goal of maximizing profits through energy trading. Two of Power TAC's main goals are to help researchers understand (1) the dynamics of customer and retailer decision-making, as well as (2) the robustness of proposed market designs. This paper contributes to the former by introducing TACTEX'13, the champion agent from the Power TAC 2013 competition. TACTEX is a complete, fully implemented agent that learns and adapts to the environment in which it operates, by heavily relying on online reinforcement learning and prediction methods. This paper details the implementation of TACTEX's constituent components and evaluates the performance of TACTEX and the contributions of each of these components through analysis of the competition and subsequent controlled experiments.

2 Power TAC Game Description

We start with an overview of the main elements of the Power TAC simulation. For more details, see (Ketter, Peters, and Collins 2013) and the full game specification (available at `http://www.powertac.org`). Power TAC models a competitive retail power market in a smart-grid environment of a medium-sized city. The simulation proceeds in a series of discrete timeslots, each representing one hour in the simulated world. A typical simulation runs for approximately 60 simulated days, and takes about two hours. Figure 1 shows the structure of the Power TAC simulation environment. At a high level, autonomous broker agents compete with each other by acting in three markets: (1) a wholesale market, in which energy is traded with traditional generation companies, (2) a tariff market, in which energy is traded with consumers, as well as distributed renewable energy producers, and (3) a balancing market, which serves to ensure that supply and demand are balanced at all times, and which determines the fees paid by brokers whenever they contribute to the total imbalance.

More specifically, local renewable energy producers (solar, wind) generate electric energy that is transmitted to the grid whenever weather conditions (wind, sun) allow for electricity generation. Power TAC uses real weather data and forecasts from different geographical zones. Local consumers such as office buildings, residential houses, and hospitals, consume energy according to their needs, and based on weather conditions and calendar factors such as day of week and hour of day. All customers are equipped with smart-meters, so consumption and production are reported every hour. Autonomous broker agents compete with each other on gaining market share and maximizing profits by trading electricity. Brokers interact with local producers and consumers through a retail market, called *tariff market*, by publishing tariff contracts for energy consumption/production that may include usage and per-day charges, fixed and varying prices, signup bonuses, and early withdrawal penalties. Customers have ranges of preferences over tariff terms and they subscribe to tariffs they find attractive. Some factors that affect preferences are expected costs of tariffs, and the expected inconvenience of shifting consumptions to times when energy is cheaper due to variable-rate tariffs. Some customers represent whole populations (e.g. a village of 30,000 people) and can subscribe subsets of their populations to different tariffs. Brokers may publish one or more tariffs once every 6 hours, 4 times per day.

In addition to the tariff market Power TAC models a traditional *wholesale market* in which brokers can buy or sell energy for future delivery. The wholesale market is a day-ahead market modeled based on real-world North American and European markets. At any given time, brokers may place orders in the wholesale market to buy or sell power in parallel, independent 24 double-auctions, where each auction is for power to be delivered in one of the next 24 hours (timeslots). For instance, a broker may procure energy in the wholesale market to satisfy the predicted demand of its portfolio. Other main participants in the wholesale market are large generation companies, that typically sell power and simulate utility-scale power suppliers.

On the electricity grid, supply and demand must be balanced at all times. Power TAC assumes that any imbalance is resolved by transmitting energy to/from the national grid, where prices are determined in a *balancing market*. Typically, a broker has a strong incentive to maintain a supply-demand balance, since the prices in the balancing market are much less attractive then the prices in the two other markets. Finally, a distribution utility charges the broker a fixed price for each kWh of energy distributed to the broker's portfolio.

The state of the game is rich and high-dimensional: it includes the set of all active tariffs and customer subscriptions, the wholesale market deliveries and orders of all brokers for the next 24 hours, the current energy consumption of all customers, the current weather and weather forecast, the current time, and the bank balance of all brokers. The game state is partially observable to brokers. For instance, brokers sees all published tariffs in the tariff market but they only know the customer subscriptions for their own tariffs. Similarly, when an auction finishes in the wholesale market, brokers only see the clearing price of the auction and a list of uncleared orders, but they do not know the list of cleared orders, or the future deliveries of other brokers. The action space of brokers is also high-dimensional. For instance, tariff publications can include up to $7 \times 24 = 168$ hourly energy prices, and wholesale market actions can include up to 24 parallel limit orders of the form bid(energy-amount,limit-price).

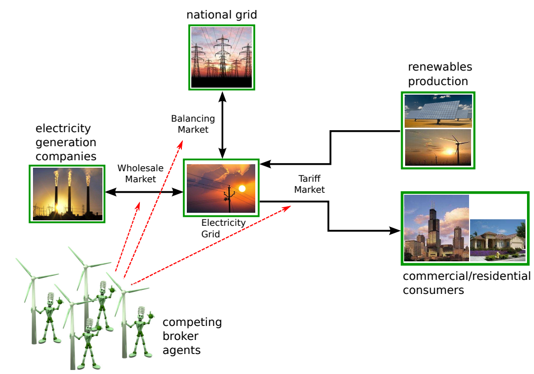

Figure 1: High-level structure of the Power TAC game

3 The TacTex'13 Agent

TACTEX is a utility-maximizing broker agent that operates simultaneously in multiple markets. TACTEX's utility measure is the cash amount in its bank account, called *cash position*. At each timeslot, TACTEX executes zero or more actions in both the tariff market and the wholesale market. The executed actions are those that are predicted to maximize its expected long-term utility. In the tariff market, the actions considered by TACTEX are consumption-tariff publications, while in the wholesale market the considered actions are bids and asks, to procure and sell energy respectively.

To maximize its utility, TACTEX must simultaneously optimize its income and costs and find a long-term profit-maximizing combination of (1) energy-selling prices (denoted p), (2) total energy demand of its customers (denoted D, controllable by how many customers it agrees to serve), and (3) energy-procurement costs (denoted C) for satisfying this demand. Fully optimizing this combined decision-making problem is intractable; therefore TACTEX approximates its solution. Let t be some future timeslot, and let p_t, D_t, and C_t be TACTEX's timeslot-specific published energy-price, customers' energy-demand and energy-procurement costs, respectively. Let $u_t(D_t, C_t, p_t) = D_t \times (p_t - C_t)$ be the utility (i.e. profit) contribution at time t. Let $\hat{D}_t, \hat{C}_t, \hat{p}_t$ be the current predictions of D_t, C_t, p_t. Let $\mathcal{A} := A_D \cup A_C \cup p$ be the set of available actions, where here A_D and p are tariffs and price publications, and A_C are wholesale market bids. Let $A_t \subset \mathcal{A}$ be the subset of actions that is taken at timeslot t. TACTEX approximates a solution to the following interdependent optimization problems (using '$+i$' to denote 'i timeslots into the future'):

1. Optimize costs given predicted demand:

$$\underset{\{A_{Ct}\}_{t=+1}^{+T}}{\arg\max} \sum_{t=+1}^{+T} E[u_t(\hat{D}_t, C_t, \hat{p}_t)] \tag{1}$$

2. Optimize demand and selling prices given predicted costs:

$$\underset{\{A_{Dt}, p_t\}_{t=+1}^{+T}}{\arg\max} \sum_{t=+1}^{+T} E[u_t(D_t, \hat{C}_t, p_t)] \tag{2}$$

Thus, instead of optimizing over all possible combinations, we separately optimize demand and costs, each conditioned on the current estimate of the other. Each of the two interdependent optimizations perform local improvement steps, however the gain with respect to global optimization is a reduction of the search complexity from multiplicative to additive. The two optimization problems defined by Equations 1 and 2 are still individually intractable. Their solutions are approximated by TACTEX's two constituent components: its wholesale market strategy (Section 3.1), and its tariff market strategy (Section 3.2).

3.1 Wholesale Market Strategy

In the wholesale market, TACTEX optimizes the costs C_t of procuring the energy needed for satisfying the predicted demand $\hat{D}_t$ resulting from selling-prices $\hat{p}_t$. $\hat{p}_t$ is assumed to be the currently published prices; $\hat{D}_t$ is predicted from $\hat{p}_t$ similarly to the demand prediction described in Section 3.2.

To Minimize the energy costs C_t, TACTEX needs to (1) minimize the rates for which it procures energy in the wholesale market, and (2) minimize its imbalance costs, by satisfying the future demand as accurately as possible. To do the latter, it must (2.1) have accurate predictions of future demand, and (2.2) be able to procure all the energy predicted to be demanded. The actions that affect the energy cost for a *target timeslot* t_{tar} are the 24 bidding (or not-bidding) actions in each of the 24 preceding timeslots, $(t_{tar} - 24)$–$(t_{tar} - 1)$, which thus comprise a sequential bidding process with 24 steps. Thus, at each timeslot t , TACTEX executes, respectively, steps $1, 2, \ldots, 24$ of 24 independent bidding processes for timeslots $t + 24, \ldots, t + 1$.

TACTEX's wholesale market bidding strategy uses a modified version of Tesauro's bidding algorithm (Tesauro and Bredin 2002). We model the sequential bidding process as a Markov Decision Process (MDP) (Puterman 1994) in a specific way that allows for computational efficiency, and more importantly in the competitive environment that TACTEX operates in, it allows for high reuse of data, and thus quick online learning with little data. TACTEX's MDP is defined next, followed by the rationale behind its design:

- **States:** $s \in \{0, 1, \ldots, 24, success\}$, $s_0 := 24$
- **Actions:** limit-price $\in \mathbb{R}$
- **Transition:** 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 $p_{cleared}(s, \text{limit-price})$ is initially unknown.
- **Reward:** In state $s = 0$, the reward is the balancing-price per energy unit. In states $s \in \{1, \ldots, 24\}$, the reward is 0. In state $success$, the reward is the limit-price of the successful bid. Both balancing-price and limit-price are taken as negative, so maximizing the reward results in minimizing costs. balancing-price is initially unknown.
- **Terminal States:** $\{0, success\}$

In a sequential bidding process for a target timeslot, the broker actions are bids of the form bid(energy-amount,limit-price). Tesauro's bidding MDP uses these actions as the MDP actions. However, in TACTEX's MDP model energy-amount is not part of the decision making; it is always set to the difference between predicted demand and the energy that is already procured for the target timeslot. The solution to our MDP is a sequential bidding strategy that minimizes the expected energy unit-cost for the next fraction of the procured amount. Note that there is a transition to a terminal state $success$ even in cases where the bid is partially cleared. One implication of excluding energy-amount from the MDP's state and action representations is that every sequential bidding process executes over the *same sequence of states*. As seen next, this allows for computational and data efficiency.

Since the MDP is acyclic (linear), solving it requires one back-sweep, starting from state 0 back to state 24, applying the following backup operator to compute a value function:

$$V(s) = \begin{cases} \text{balancing-price} & \text{if } s = 0 \\ min_{\text{limit-price}}\{p_{cleared} \times \text{limit-price} + \\ (1 - p_{cleared}) \times V(s-1)\} & \text{if } 1 \leq s \leq 24 \end{cases}$$

The MDP's solution determines an optimal limit-price for each of the 24 states. Using our MDP model, TACTEX is always in states $1, \ldots, 24$ of 24 concurrent bidding processes. Therefore, TACTEX solves the MDP once per timeslot, and submits the 24 optimal limit-prices to the 24 auctions.

Before solving this MDP, TACTEX needs to learn the MDP's unknown quantities, namely the expected balancing-price at $s = 0$ and the transition function $p_{cleared}$. TACTEX learns the transition function from past data by recording, for each state $s \in \{1, \ldots, 24\}$, the wholesale trades executed in s into a set $\mathcal{P}_s$. Each trade has the form (clearing-price,cleared-energy-amount). The set $\mathcal{P}_s$ is treated as a non-parametric density estimation and a transition probability is estimated from it as $p_{cleared}(s, \text{limit-price}) := \frac{\sum_{tr \in trades[s], tr.\text{clearing-price} < \text{limit-price}} tr.\text{cleared-energy-amount}}{\sum_{tr \in trades[s]} tr.\text{cleared-energy-amount}}$. To estimate the mean balancing-price, TACTEX similarly maintains a set $\mathcal{P}_0$ of past balancing data. Since every bidding MDP executes over the same sequence of states $s \in \{0, \ldots, 24\}$, every trade executed in state s can be used by all future bidding processes as a part of $\mathcal{P}_s$. Thus, our state representation allows TACTEX to efficiently reuse data and thus speed-up learning. Clearly, our state representation relies on the assumption that time-to-target-timeslot is a dominant feature in determining the transition function, i.e. the distribution of auction closing prices. Were that not the case, other features would need to be added to the MDP's state.

TACTEX's bidding strategy is summarized in Algorithm 1 which is TACTEX's main routine in the wholesale market, executed at every timeslot. It computes the needed energy for the coming 24 timeslots using demand-predictions (line 1), then adds the previous timeslot's wholesale market trades and balancing information to the $\mathcal{P}_s$ sets (line 2). If not enough (specifically fewer than 6) trades were recorded for each state, a randomized bidding strategy is executed, otherwise the MDP-based bidding strategy is executed (lines 3-7). The number 6 was chosen to trade off quick learning with reasonable density estimations.

Algorithm 1 Online RL Wholesale Market Strategy

1: neededEnergy[1 . . . 24] = ComputeNeededEnergy()
2: densities[0 . . . 24] $\leftarrow$ AddRecentTradesAndBalancing()
3: **if** HasEnoughData(densities) **then**
4:     limitPrices[1 . . . 24] = SolveMDP(densities)
5: **else**
6:     limitPrices[1 . . . 24] = RandomizedBiddingPolicy()
7: SubmitBids(neededEnergy[1 . . . 24], limitprices[1 . . . 24])

To summarize, TACTEX starts a game with no data and learns to bid online, while acting. Its estimates are refined during the game as it collects more data. At each timeslot, it solves the MDP with all the data collected so far. The result is an online reinforcement learning (RL) bidding algorithm that allows TACTEX to adapt and optimize its bidding strategy to each game's specific market conditions.

3.2 Tariff Market Strategy

In the tariff market, TACTEX optimizes future demands D_t and selling-prices p_t given the predicted energy costs $\hat{C}_t$. Algorithm 2 is TACTEX's main routine in the tariff market, executed at every tariff-publication timeslot. It starts by generating a set of 100 fixed-rate candidate tariffs, with rates that are equally spaced in a range that contains the current best (lowest) published rates (line 1). Next, EstimateUtility() predicts the expected long-term utility of each candidate tariff-publication action, and the action with the highest predicted value is executed (lines 2-6).

Algorithm 2 Utility-based Tariff Market Strategy

1: candidateTariffs $\leftarrow$ GenerateCandidateTariffs()
2: **for** tariff in candidateTariffs **do**
3:     utilities[tariff] $\leftarrow$ EstimateUtility(tariff)
4: bestTariff, bestUtility $\leftarrow$ FindMax(utilities)
5: **if** bestUtility $>$ EstimateUtility(no-op) **then**
6:     PublishTariff(bestTariff)

Due to the high-dimensionality of the state-space, Algorithm 2 is implemented as a *lookahead policy* (Powell 2011), where candidate actions' long-term utilities are estimated using a lookahead search (also called Monte Carlo search) over some future horizon starting at the current state. The length of the horizon over which utility is estimated is one week ($7 \times 24 = 168$ timeslots), chosen as a trade-off between shorter horizons, which might not capture weekly consumption/production patterns, and longer horizons which present higher uncertainty and require more computation.

Using a lookahead policy aims at reducing the complexity of searching in high-dimensional *state-spaces*. To reduce the complexity of searching over high-dimensional *action-spaces*, or on top of that, over sequences of subsets of high-dimensional actions, TACTEX further approximates the solution to the search problem in several ways. First, TACTEX searches over *single* actions rather than over *subsets* of actions, by considering only a single tariff-publication at a time. Second, instead of running a tree-search over sequences of future actions, TACTEX estimates the current action's utility assuming no other tariff actions are going to be taken during the lookahead horizon. Third, TACTEX searches solely over one type of action, namely fixed-rate tariff publications, and therefore optimizes only one selling-price rather than a separate price for each future timeslot.

EstimateUtility() works by estimating long-term income and costs after publishing a candidate tariff, as described in Algorithm 3. At the core of EstimateUtility() lies the problem of estimating the demand resulting from a tariff publication, which in turn is decomposed into the two problems of predicting (1) the resulting customer migration between tariffs (line 3 and Algorithm 4), and (2) the demand of each of the customers over the lookahead horizon (line 4). The latter is addressed by maintaining records with average past demand for each customer, in each of the 168 weekly timeslots. Predicting energy costs (line 5) is addressed similarly by maintaining a record of average past costs in each of the 168 weekly timeslots. Using the information gathered in lines 3-5, the total utility is computed by summing over all customer-tariff pairs (line 6-13). A one-time tariff publication fee is added to the estimated costs (line 12).

Algorithm 3 EstimateUtility(tariff)

1: totalIncome $\leftarrow$ 0
2: totalCosts $\leftarrow$ 0
3: subs $\leftarrow$ PredictChangeInSubscriptions(tariff)
4: demandProfiles $\leftarrow$ PredictCustomerDemandProfiles()
5: costs[1 . . . 168] $\leftarrow$ PredictEnergyCosts()
6: **for** cust in customers **do**
7:     **for** tariff in $\{$tariff $\cup$ existingTariffs$\}$ **do**
8:         n $\leftarrow$ subs[cust,tariff]
9:         demand[1 . . . 168] $\leftarrow$ n $\times$ demandProfiles[cust]
10:         totalIncome += $\sum_{i=1}^{168}$ demand[i] $\times$ tariff.rate()
11:         totalCosts += $\sum_{i=1}^{168}$ demand[i] $\times$ costs[i]
12: totalCosts += pubFee(tariff)
13: **return** totalIncome - totalCosts

Algorithm 4 describes how TACTEX predicts the changes in subscriptions as a result of a new tariff publication. TACTEX focuses on customers that represent whole populations and can subscribe subsets of their population to different tariffs. TACTEX predicts the change in subscriptions separately for each of these customers (line 2). It multiplies the predicted weekly demand of a single member of the population (line 3) with a tariff's rate to compute the expected weekly charge for a single member under this tariff (line 6). The pairs $\langle charge, subs \rangle$ of existing-tariffs' expected charges and subscribed-populations are used as a training set for a supervised learning algorithm, specifically Locally Weighted Linear Regression (LWR), that predicts the subscribed-population size for a candidate tariff based on its expected charge (lines 8-12). LWR (see, e.g. (Atkeson, Moore, and Schaal 1997)) was chosen since, being non-parametric, it requires very minimal assumptions about the representation of the predicted function (the customer preference function). Since new subscribers to the candidate tariff must migrate from other tariffs (published by either TACTEX or its competitors), predicted subscriptions are scaled down to sum to the customer's population (line 13).

Algorithm 4 PredictChangeInSubscriptions(tariff)

```
allTariffs ← {tariff ∪ existingTariffs}
for cust in customers do
    demand[1 . . . 168] ← PredictDemandProfile(cust)
    charge2subs ← {}
    for t in existingTariffs do
        charge ← ExpectedTariffCharge(demand, t)
        subs ← GetNumCurrentSubscriptions(cust, t)
        charge2subs ← charge2subs ∪ ⟨charge, subs⟩
    charge ← ExpectedTariffCharge(demand, tariff)
    trainingSet ← charge2subs
    subs ← PredictWithLWR(trainingSet, charge)
    charge2subs ← charge2subs ∪ ⟨charge, subs⟩
    charge2subs ← Normalize(charge2subs)
    for t in allTariffs do
        predSubs[cust,t] ← ExtractSubscriptions(charge2subs,t)
return predSubs
```

4 Results

This section examines the success of TacTex through analysis of the competition and controlled experiments.

4.1 Power TAC 2013 Finals Analysis

The Power TAC 2013 finals were held in conjunction with the AAAI'13 conference. The qualifying competitors were 7 brokers developed by research groups from Europe and the USA. The competition included all possible combinations of 2-broker and 4-broker games (21 and 35 games respectively), and 4 7-broker games. Table 1 shows the final cumulative scores in each of the game sizes. In the 2-agent games TacTex won all of its 6 games. In the 4-agent games, TacTex won 15 out of the 16 games it completed successfully (TacTex got disconnected from 4 games due to technical issues with the infrastructure we used). TacTex did not win the 7-agent games despite having the largest volume of customers. Next, we analyze these results.

Table 1: Results of the Power TAC 2013 finals

Broker	7-broker	4-broker	2-broker	Total (not normalized)
TacTex	-705248	13493825	17853189	30641766
cwiBroker	647400	12197772	13476434	26321606
MLLBroker	8533	3305131	9482400	12796064
CrocodileAgent	-361939	1592764	7105236	8336061
AstonTAC	345300	5977354	5484780	11807435
Mertacor	-621040	1279380	4919087	5577427
INAOEBroker02	-76112159	-497131383	-70255037	-643498580

Figure 2 shows averages of the main components of the brokers' cash flow, for each of the game sizes.[1] Brokers are ordered based on their final ranking in the competition, from left to right. For each broker, the bars show (from left to right) its average (1) profit (2) income from consumption tariff subscribers (3) tariff publication fees (proportional the number of tariffs published) (4) wholesale market costs (5) balancing costs, and (6) energy distribution costs (proportional to the amount of traded energy).

At a high level, TacTex's wholesale market strategy and tariff market strategy were responsible for TacTex's success in the finals. The wholesale market strategy maintained low-costs, while the tariff market strategy balanced its offered tariff prices with the resulting predicted demand to optimize profits given the costs achieved by the wholesale strategy. More specifically, in the 2-agent games TacTex

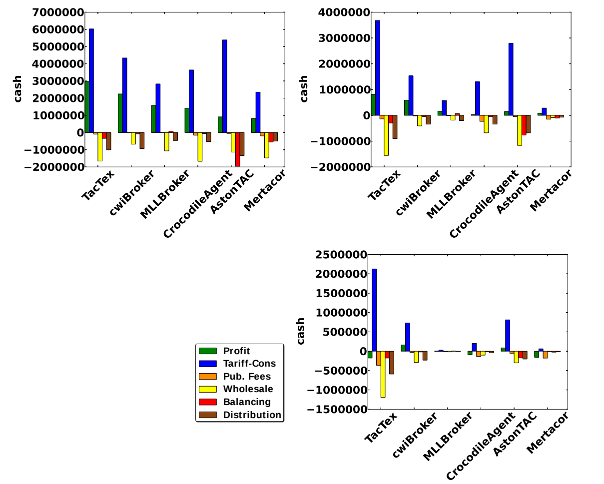

Figure 2: Power TAC 2013 Finals: avg. income/costs in 2-agent (top-left), 4-agent (top-right), and 7-agent games (bottom-right)

made 32.4% and 88.2% more profits than the 2nd (cwi) and 3rd (MLL) place brokers while maintaining similar levels of income-to-costs ratio (1.97), compared to cwi's (2.07) and MLL's (2.26). In Power TAC's wholesale market, energy unit-cost is typically an increasing function of the procured amount. Despite that, TacTex sold 50.5% and 72.5% more energy then cwi and MLL with a competitive cost-per-kWh (4.4 cents/kWh) compared to cwi's and MLL's (4.6, 3.1 cents/kWh)[2]. It can be seen that the majority of TacTex's costs were spent on wholesale market procurement and (non-controllable) distribution fees. Therefore, TacTex's low cost-per-kWh is attributed to its wholesale market strategy. At the same time, given these costs, its tariff market strategy published tariffs at an average rate that is slightly lower than cwi's and slightly higher than MLL's (8.8, vs 9.5 and 7.1 cents/kWh), which resulted in 39.0% and 113.6% more income compared to cwi and MLL. In the 4-agent games, TacTex traded 9% less energy comparing to the 2-agent games, while maintaining similar average wholesale market costs. Due to the stronger competition, TacTex's income decreased by 61%, since its tariff market strategy recognized it had to reduce prices (by 66.6%) to maximize its profits. TacTex's profits (and income) were higher by 38.1% (139.9%) and 404.5% (542.2%) compared to cwi's and MLL's, while its income-to-cost ratio decreased to 1.28 compared to 1.62 and 1.39 of cwi and MLL. In the 7-agent games, TacTex's tariff strategy had to lower prices further, but also recognized a stopping point beyond which it did not decrease rates. However, due to an underestimation of the predicted costs, TacTex ended up with losses despite having large customer volume and income.

4.2 Controlled Experiments

In this section we perform controlled experiments to identify the contribution of each of TacTex's major components. To do that, we generate test agents by disabling components of TacTex and comparing the resulting performance. Specifically, agent U9_MDP_LWR is the full TacTex agent. Agent

[1] We excluded INAOEBroker; its large negative scores, caused by large tariff-publication fees, affected the readability of the plots.

[2] Not shown in the figure due to space constraints.

Table 3: Ablation analysis using 3 finalist agents.

Broker	Cash	Broker	Cash	Broker	Cash	Broker	Cash	Broker	Cash
cwiBroker	340.9 (8.4)	cwiBroker	315.4 (9.3)	cwiBroker	316.2 (9.1)	**U9_MDP**	**389.9 (13.3)**	**U9_MDP_LWR**	**350.8 (13.3)**
Mertacor	-276.2 (40.2)	**U1**	**135.3 (12.3)**	**U9**	**182.8 (12.4)**	cwiBroker	138.3 (8.7)	cwiBroker	132.4 (9.0)
CrocodileAgent	-287.1 (14.5)	CrocodileAgent	-372.1 (17.0)	CrocodileAgent	-338.2 (17.0)	CrocodileAgent	-333.3 (17.0)	CrocodileAgent	-336.9 (17.3)
B	**-334.6 (8.0)**	Mertacor	-485.5 (28.1)	Mertacor	-476.6 (28.6)	Mertacor	-494.1 (29.6)	Mertacor	-566.1 (26.8)

U9_MDP is generated from U9_MDP_LWR by removing the LWR-based customer-subscriptions prediction component and replacing it with linear interpolation and conservative extrapolation. Agent U9 is generated from U9_MDP by disabling the MDP-based wholesale market strategy and replacing it with a baseline, randomized strategy that starts by trying lower buying prices and increasing them as time gets closer to target timeslot. Agent U1 was generated from U9 by publishing 1, instead of 9, initial sample tariffs for probing customer tariff subscriptions, used by Algorithm 4. Finally, a baseline agent B was generated from U1 by disabling the tariff-market strategy (Algorithm 2), and replacing it with a strategy that reacts to competitor tariffs by publishing slightly better rates.

Table 2: Round-Robin ablation analysis.

	B	U1	U9_MDP
U9_MDP_LWR	1278.3 (43.2)	708.9 (35.6)	34.2 (23.2)
U9_MDP	966.4 (40.5)	592.6 (22.2)	
U1	547.4 (27.7))		

We compared the above agents in two groups of experiments. The first group is a 2-agent round-robin tournament between U9_MDP_LWR, U9_MDP, U1 and B. The second group compared the performance of all versions in 4-agent games against a fixed set of opponents, composed of the 3 finalist agent binaries that are currently available to us: cwiBroker, CrocodileAgent and Mertacor. In all of our experiments, each given combination of agents was tested over a fixed set of 200 full games. Each game takes about 2 hours of real-time (about 60 days of simulated time), and was generated by loading a set of random-number seeds that initialize the random number generators of the simulation, and a weather data file that completely determines the simulated weather. We note that even after loading weather and seeds, there is still some randomness of unknown source in the simulation. Each weather file contains around 3 months of real-world weather, recorded in the default location simulated by Power TAC. We used 8 different weather files (each file used by 25 out of the 200 games), using the recording start dates of January, April, July, October of 2009 and 2010, thus covering a period of 2 years. The results of the first group of experiments are reported in Table 2. Each entry in the table is the the mean score-difference (in 1000s) over the set of 200 games. The results of the second group of experiments is reported in Table 3. Each of the 5 two-column sub-tables shows the results when playing one of our agent versions against the 3 finalist agents over the set of 200 games. Each entry shows the average score of each agent, and rows are ordered by ranking. In both groups, adding the tariff market strategy and the wholesale market strategies resulted in significant improvements. Adding the LWR-based prediction (U9_MDP_LWR) seems to be beneficial only for 2-agent games, possibly since its less conservative extrapolations work better with small number of competitors.

5 Related Work

Since Power TAC is a new domain there is not much published work on other broker approaches. SELF (Peters et al. 2013) uses the sarsa RL algorithm for selecting tariff market actions. It is designed to choose one of 6 tariffs actions and therefore is less flexible then TacTex's tariff market strategy, which is not constrained to a finite set of actions. The AstonTAC agent uses a different MDP model for wholesale bidding (Kuate et al. 2013). It assumes an underlying discrete model for wholesale clearing prices (HMM), where 20 possible states are built offline from a game's bootstrap data. Our MDP does not assume an underlying model of the market, but rather uses a more flexible, non-parametric model of clearing prices at every state. Furthermore, TacTex uses a different state representation, designed to allow high reuse of data and computation, and therefore fast learning. Our bidding algorithm is a modified version of Tesauro's bidding algorithm for double-auctions (Tesauro and Bredin 2002), as explained in Section 3.2. Earlier approaches to broker development (Reddy and Veloso 2011a; 2011b) worked under a more limited setup that did not include wholesale trading, and assumed fixed customer consumption instead of the variable daily load profile of Power TAC customers. Utility-based approaches to trading agents were presented in the prior trading agent competitions, however the game setups and the problems they solved, and consequently the methods used, were different than TacTex's utility-based approach in the Power TAC domain. In TAC-travel, decision theoretic bidding using Monte-Carlo estimation of the clearing price distribution was used for one-sided auctions (Stone et al. 2003). In TAC Supply Chain Management, (Pardoe 2011) used an interdependent optimization of supply and demand, but trading was done through requests for quotes, rather than through a tariff market and a wholesale market as in the Power TAC setup. In past TAC competitions, other proposed approaches to agent design included a game theoretic analysis of the economy (Kiekintveld, Vorobeychik, and Wellman 2006) and fuzzy reasoning (He et al. 2005). Interdependent optimization in a different but related context is executed by the simulated customers in Power TAC (Reddy and Veloso 2012), to simultaneously optimize tariff subscriptions and consumption profiles.

6 Conclusions

This paper introduced TacTex, the champion power trading agent from the Power TAC 2013 finals. This paper describes the complex decision making problem faced by a broker agent, and details and evaluates the approximate solution implemented by TacTex. Future research directions include investigating the usage and optimization of different types of tariffs and energy balancing methods, as well as their overall impact on the smart grid.

This work has taken place in the Learning Agents Research Group (LARG) at the Artificial Intelligence Laboratory, The University of Texas at Austin. LARG research is supported in part by grants from the National Science Foundation (CNS-1330072, CNS-1305287), ONR (21C184-01), and Yujin Robot.

References

Atkeson, C. G.; Moore, A. W.; and Schaal, S. 1997. locally weighted learning.

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

European Commission. 2011. *Communication: Energy Roadmap 2050*.

He, M.; Rogers, A.; David, E.; and Jennings, N. R. 2005. Designing and evaluating an adaptive trading agent for supply chain management applications. In Poutré, H. L.; Sadeh, N.; and Sverker, J., eds., *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*. Springer. 35–42. Event Dates: Auguest 2005.

Ketter, W.; Collins, J.; and Reddy, P. 2013. Power TAC: A competitive economic simulation of the smart grid. *Energy Economics* 39(0):262 – 270.

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

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

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

Pardoe, D. M. 2011. *Adaptive Trading Agent Strategies Using Market Experience*. Ph.D. Dissertation.

Peters, M.; Ketter, W.; Saar-Tsechansky, M.; and Collins, J. 2013. A reinforcement learning approach to autonomous decision-making in smart electricity markets. *Machine Learning* 92(1):5–39.

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

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

Ramchurn, S.; Vytelingum, P.; Rogers, A.; and Jennings, N. R. 2012. Putting the "smarts" into the smart grid: A grand challenge for artificial intelligence. *Communications of the ACM* 55(4):86–97.

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

Reddy, P. P., and Veloso, M. M. 2011b. Strategy learning for autonomous agents in smart grid markets. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Two*, 1446–1451. AAAI Press.

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

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

Tesauro, G., and Bredin, J. L. 2002. 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, 591–598. New York, NY, USA: ACM.

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