MultiCraft

International Journal of Engineering, Science and Technology
Vol. 2, No. 3, 2010, pp. 45-59

INTERNATIONAL JOURNAL OF ENGINEERING, SCIENCE AND TECHNOLOGY
www.ijest-ng.com

Intelligent Agents under Collaborative Control in Emerging Power Systems

H.F. Wedde[1], S. Lehnhoff[1*], C. Rehtanz[2], O. Krause[2]

[1]Department of Computer Science, Technical University of Dortmund, GERMANY
[2] Department of Electrical Engineering, Technical University of Dortmund, GERMANY
[*]Corresponding Author: e-mail: sebastian.lehnhoff@tu-dortmund.de, Tel +49-231-755-2768, Fax+49-231-755-5323

Abstract

In the world of liberalized power markets traditional power management concepts have come to their limits. Optimal pricing can no longer be achieved, e.g. for very short-time needs across grids. Power line overload and grid stability, increasingly resulting in regional or even global black-outs, are at stake. With the highly desirable expansion of renewable energy production these challenges are experienced in quite an amplified way: We argue that for this emergent technology the traditional top-down and long-term power management is obsolete, due to the wide dispersion and high unpredictability of wind and solar-based power facilities. In the DECENT[1] R&D initiative we developed a multi-level, bottom-up solution where autonomous collaborative software agents negotiate available energy quantities and needs on behalf of consumer and producer groups (the DEZENT algorithm). We operate within very short time intervals of assumedly constant demand and supply, in our case periods of 0.5sec (switching delay for a light bulb). The solution has proven to be secure against a relevant variety of malicious attacks. Within this time interval we are also able to manage the coordinated power distribution, and achieve grid stability. In this paper the main contribution is to make the negotiation strategies themselves adaptive across periods: We derive the dynamic distributed learning algorithm DECOLEARN from Reinforcement Learning principles for providing the agents with collaborative intelligence and at the same time proving substantially superior to conventional (static) procedures. We report briefly on our extensive comparative simulation experiments.

Keywords: Distributed Energy Management, Reinforcement Learning, Multi-agent Systems, Smart Grids

1. Novel Perspectives and Challenges through Regenerative Electric Energy

In the world of liberalized power markets traditional power management concepts have come to their limits. Optimal pricing can no longer be achieved, e.g. for very short-time needs across grids. Avoiding power line overload and maintaining grid stability (increasingly resulting in regional or even global black-outs) are at stake. With the both ecologically and economically desirable oncome of renewable energy production these challenges exhibit quite a fundamental quality, making a strong plea for these negotiation and management problem areas to be *integrated* into an adequate novel approach.

1.1 Distributed Agent Negotiation of Regenerative Power: In Southern Germany approx. 180 cities and townships are already able to cover their needs through wind craft, solar power, or through small and mid-size *block heat & power plants (BHPPs)* driven by seed oil. Even entire regions in Germany and Austria are close to completely support themselves with clean energy. There are similar efforts and projects within the European Community and world-wide. This will cause *a paradigm shift:* The traditional power production, distribution, and management are based on long-term negotiations and on a *centralized top-down balancing of needs and supply,* entailing network stability problems (rather frequent black-outs) as much as high environmental and financial costs for reserve facilities (1-min reserve and below) running permanently on a 20% efficiency level.

[1] This work is part of the DECENT R&D initiative. It is partially funded by Deutsche Forschungsgemeinschaft (German Research Foundation, DFG), under contracts WE 2816/4-1 and HA 937/32-1, and by E.ON (contract no. 4500014170).

In turn, regenerative energy production is based on widely dispersed small or mid-size facilities. Balancing production and consumption are organized *bottom-up,* coordinated between local sources, and stepwise up between ever larger regions. To a large extent, producers are also consumers, and vice versa. In the best common interest of the users involved, and at the same time respecting the responsibility of the local and regional customers (producers as much as consumers), these should plan for largely covering their normal needs themselves yet count for balancing power from higher levels in (unexpected) peak times. The management itself can only be done *under decentralized control,* due to the problem complexity which renders a centralized management obsolete, even under support of very large computer systems. On the other hand, every single source is a potential reserve facility (yielding a comparably high fault tolerance). As long as the needed power could be negotiated and distributed *in due time* no dedicated reserve power stations need to be kept, contrary to the traditional practice. Since the availability of wind or solar power is rather unpredictable balancing between such sources could be backed up by BHPPs, or by novel batteries featuring extremely high capacities and very high power gradients while being charged or discharged (Burnett *et al.*, 2006).

1.2 Negotiation Periods and Distributed Intelligence: Under the highly unpredictable fluctuation of solar and wind energy transformation facilities – and generally of local production and consumption –, at the same time respecting the authority and interests of local customers, our key objective is for the distributed negotiations to take place during time intervals of assumedly constant need and supply. In fact, we have chosen the smallest reasonable size, namely 0.5 sec (the delay between turning a light switch and the bulb being in full operation). This assumption is acceptable, due to the laxity of electric devices, actors, and procedures. These intervals are termed *periods,* designed for negotiations of distributed agents. Even for software agents this creates rather rigid real-time deadlines.

A human actor will set up, or modify, a negotiation strategy on his level of granularity, thus even a strategy which is meant to be dynamic on this scale would appear mostly static in the view of the agents! Furthermore, even 1 min (120 periods!) is too short in the human perspective for readjusting strategies. To take advantage of the speed of the agents and at the same time close the gap between human actor and agent behavior we provide for the agents to intelligently revise their strategies, *for their next period*. In the absence of global information or any safe prediction basis, we derived distributed learning strategies from distributed Reinforcement Learning.

In each interval, our strategies are exponential bidding functions. While they formalize organic growth or decline they have been found most suitable for fast convergence between producer offers and consumer bids. A few of the technical features of the distributed negotiation algorithm (DEZENT) will be described in section 2. More details can e.g. be found in Wedde *et al.*, 2008.

1.3 Previous and Related Work: There has been quite a variety of work on various aspects of Computational Intelligence research in emergent power systems for many years. Detailed investigations are concerned with (optimal) pricing (Ahamed *et al.*, 2002, Anthony *et al.*, 2001, Bakirtzis *et al.*, 2006, Hommelberg *et al.*, 2007, Nanduri *et al.*, 2007, Takano *et al.*, 1996, Tellidou *et al.*, 2006 and Zhou *et al.*, 2003), either in day-ahead or spot market models, blackboard-based auction systems, or other global tools, such as global data bases. Let alone the restriction to negotiations none of the solutions is really scalable to large systems integrating widely dispersed regenerative sources. In turn, the customers in DECENT carry out distributed negotiations through their agents, and neither look-ahead nor auctioning are needed (Wedde *et al.*, 2008). Grid stability is addressed in Ernst *et al.*, 2002, Ernst *et al.*, 2004, Hadidi *et al.*, 2009, Pipattanasomporn *et al.*, 2009 and Vlachogiannis *et al.*, 2004 in terms of various learning strategies where a central agency would be enabled to ensure (optimal) stability. This theme has been an open issue in practice so far, and the proposed solutions are certainly not scalable. In DECENT we pursued a completely distributed approach where stability checks are done online for each period of 0.5 sec, under distributed control. The authors in Massoud *et al.*, 2005 make plea for distributed safety control/reconfiguration in a military case study, and in Potter *et al.*, 2009 a plea for a very high-level control facility is made, as a suggestion for later realistic research. Altogether the quoted work from the other authors is restricted to singular operational problems while the solution quality certainly depends on its compatibility with all other relevant aspects.

At least for integrating regenerative facilities into existing grids it is clear from 1.1 and 1.2 that a distributed intelligent solution is to be developed which integrates all relevant operational aspects. The DECENT initiative is, to the best of our knowledge, the only comprehensive research effort here.

1.4 Organization of the Paper: After a compact formal introduction into the DEZENT algorithm in section 2, our collaborative (distributed) learning algorithm DECOLEARN is derived from adapted Reinforcement Learning strategies in section 3. In section 4, it will be broadly compared with a “static” algorithm (no adjustments for the next period, despite changing production and consumption patterns) as well as with an “ideal” algorithm. In the concluding section we also hint to future work.

2. Distributed Agent Negotiations (DEZENT Algorithm)

2.1. The Model: From the ideas presented in section 1 we will now continue into more detailed model assumptions (see also figure 1A for a typical power grid structure):

1) The consumer needs can be covered within a regional grid (0.4 kV - 10 kV voltage levels), or between grids (110 kV) through regenerative energy, with very few exceptions where a reserve capacity in the 380 kV network will be accessed (see figure 1B). The grids are assumed to have no electrical failures, in our subsequent discussion.

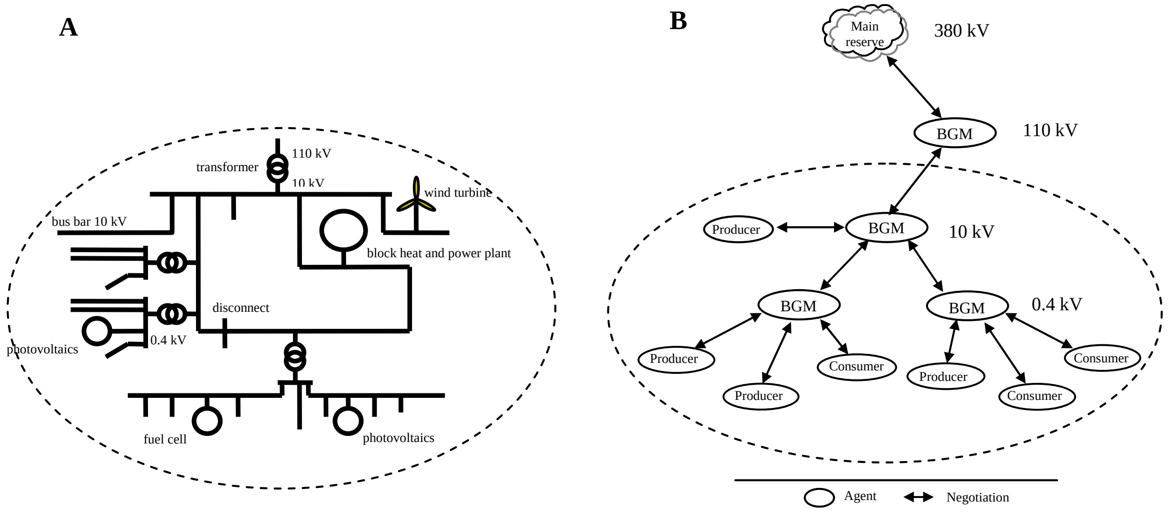


Figure 1. Power Grid and Associated Agents

2) Balancing of needs and/or of excess power may take place on different voltage levels as well as across different balancing groups on the same level (see figure 1). It will occur bottom-up from the 0.4 kV level (level 1), or between groups on the 10 kV (level 2), and 110 kV level (level 3), respectively.

3) Consumers are normally also producers, and vice versa.

4) Negotiations will be carried out through customer agents (see figure 1B) representing the human or technical actors. While the agents act autonomously their actions are, on each level, coordinated by agents called balance group managers (BGMs). These operate in parallel on each grid or inter-grid level. As an example, the grid in figure 1B contains 3 such balancing levels.

5) Negotiations are carried out within periods of 0.5 sec duration. Just during such intervals the demand and supply situations may be considered constant, a key feature of the approach. Thus such time periods while enabling stable negotiations constitute a hard deadline for all agents involved.

6) At the beginning of a period each customer agent checks its own demand and supply situation, and thus determines whether it will act as a producer (excessive power available), a consumer (additional power needed) or take no action (balanced situation).

7) Price bids and offers are limited by price frames reflecting the amortization of customer investment, maintenance, and of (low) bio-fuel costs. There are no long-term contracts, thus no discounts for large quantities (to be purchased in the form of future claims). Energy quantities are always sold and purchased for the duration of the next negotiation period (0.5 sec) only.

2.2. The Base DEZENT Algorithm: Within this model framework the main idea for setting up the distributed negotiation algorithm is as follows:

1) If $[A_k, B_k]$ is the price frame for level k ($1 \leq k \leq 3$ in figure 1B) each BGM on this level runs a coordination cycle of 10 rounds. Each round takes 1 msec.

2) After each round the BGMs check whether or not there are bids and offers "similar" enough to be matched (this is done based on a global similarity parameter), and in these cases it settles contracts between the parties. The basic idea for the BGMs is to only grant a (globally) fixed amount of X Wh (Watt hours) at a time, in a Round-Robin fashion (in Electrical Engineering, energy can be partitioned into arbitrary portions).

3) Negotiation strategies as set by a customer agent C are characterized through an opening bid $bid_C(0) \in [A_k, ½(B_k + A_k)]$, an opening offer $offer_P(0) \in [½(B_k + A_k), B_k]$, a device-specific urgency urg_0 and strategy parameters s_{1C} and t_{1P}. Furthermore, after round n; $n \in [0,9]$ the unsatisfied agents adjust their bids/offers this will be done according to:

$$bid_C(n) = -\frac{1}{e^{\frac{urg_0 \cdot n}{s_{1C}} + s_{2C}}} + B_k \tag{1}$$

$$offer_P(n) = \frac{1}{e^{\frac{urg_0 \cdot n}{t_{1P}} + t_{2P}}} + A_k \tag{2}$$

The s_{2C} and t_{2P} are determined by the opening bid ($bid_C(0) = bid_0$) or offer ($offer_P(0) = offer_0$), respectively:

$$s_{2C} = -\log(B_k - bid_0) \tag{3}$$

$$t_{2P} = -\log(offer_0 - A_k) \tag{4}$$

The exponential behavior is most suited for fast convergence of bids and offers. Figure 2 gives a pictorial impression of the negotiations between 6 consumers (ascending curves) and 5 producers (descending curves). Encircled bid/offer pairs (of similar values) and numbers correspond to the order in which contracts are closed (the similarity range in figure 2 is set to 2 ¢). On contracting either the consumer curve ends (contract 2), due to the needed quantities being smaller than offers, or the producer curve ends (contracts 3 and 4), due to offers being smaller than the needed quantities. Both curves end when needed and offered quantities match exactly (contracts 1, 5 and 6). In this example two consumers remain unsatisfied by the end of the tenth round.

4) Unsatisfied customers in a cycle on level k are moved to the next-higher level k+1. The prices frames $[A_{k+1}, B_{k+1}]$ are the result of shrinking $[A_k, B_k]$ by a fixed shrinking rate Sr according to (shrinking rates of 20% or 40% have been alternatively selected in figure 3):

$$A_k := A_0 + c(k) \tag{5}$$

$$B_k := B_0 - c(k) \tag{6}$$

with

$$c(k) := \frac{B_0 - A_0}{2} \cdot Sr \cdot k \tag{7}$$

Initial offers and bids will potentially be adjusted to fit into the shrunken frame $[A_{k+1}, B_{k+1}]$. The other strategy parameters remain unchanged. This creates a better chance for bids and offers to match. However, at the same time contracted prices are likely to be more unfavorable than on the lower level, for all parties.

5) When a consumer C_k and a producer P_k are contracted the resulting negotiated prices for the consumer ($price_{Ck}$) and for the producer ($price_{Pk}$) are calculated according to:

$$price_{C_k} := \frac{offer_{P_k}(n) + bid_{C_k}(n)}{2} + c(k) \tag{8}$$

$$price_{P_k} := \frac{offer_{P_k}(n) + bid_{C_k}(n)}{2} - c(k) \tag{9}$$

Prices are calculated as arithmetic means with additional charges of $c(k)$ (see formula (7)) being added to consumer prices and subtracted from producer fees, respectively. These surcharges guarantee that the most favorable energy prices (for consumers as well as producers) may be negotiated only on the lowest level and thus between regionally close customers.

6) Customers who are still unsatisfied after passing all grid levels (cycles) are directed to the main reserve facility. This is highly unfavorable for their business as can even be seen in the modest pricing scheme in figure 3.

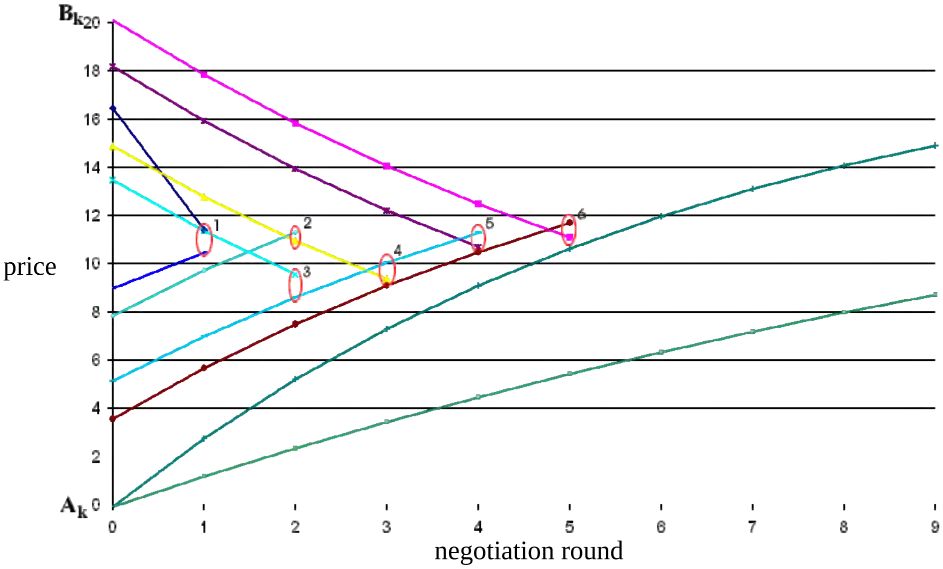


Figure 2. Contracting for Energy Quantities

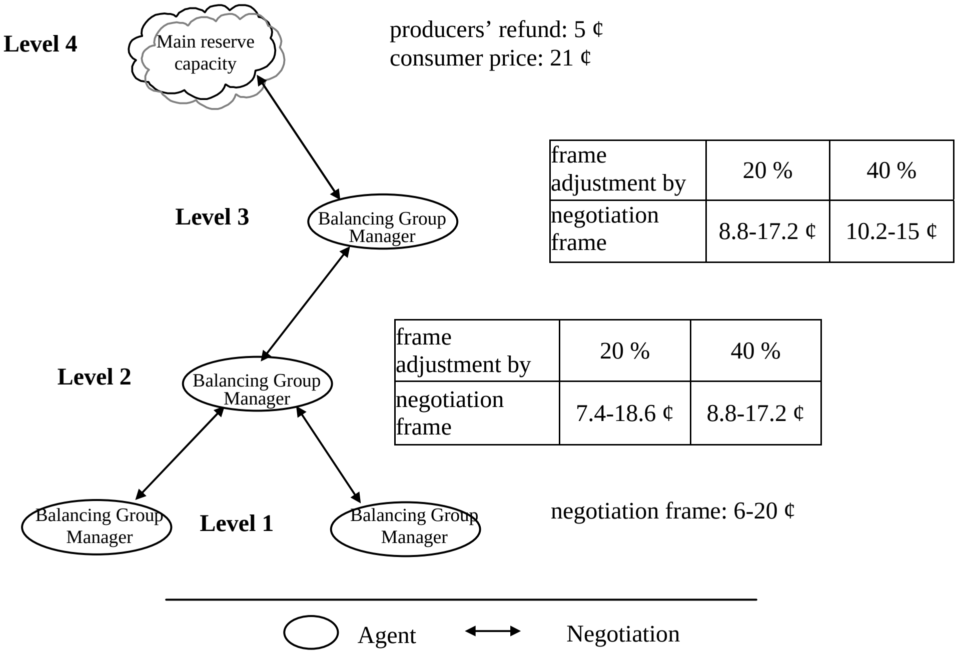
For each period, and for high numbers of customers, the deadlines previously mentioned for rounds and cycles are all met (this has been verified in extensive simulation experiments, on the high-performance Linux cluster LiDO at the Technical University of Dortmund, Germany). Since in our example – as much as in the real system – (see section 2.1) negotiations will be finalized within 40 msec this leaves more than 450 msec of the period for communication and configuring the electric power for distribution according to the negotiation results (and respecting network stability at the same time: This latter issue will be addressed in the conclusion since it is out of the scope of this paper). Also, the prices for electric energy can be kept considerably lower than under the traditional contracting between consumers and large power companies (the example in figure 3 gives a first idea. Regarding the negotiation levels please compare also figure 1B).

Also, the algorithm is robust against a large class of security attacks. Due to space limitations we refer the reader to Wedde *et al.*, 2008 for further details.

3. Periodic Reinforcement Learning in DECENT

As explained in section 2.2 negotiations are organized in cycles, and the strategies within a cycle as well as strategy adjustments (of the negotiation frames) between cycles are automated, for every period. Typically only 3-4 cycles (of 30-40 msec total duration) are needed for finalizing the negotiation process, resulting in covering the consumer needs with regenerative power, without accessing traditional (reserve) power sources (see figure3).

Between periods, i.e. every 0.5 sec (see section 2.2.1), a different form of adaptation has been established. As announced in section 1, and in face of lacking both global information and predictive insights, we have assumed a Reinforcement Learning approach Sutton *et al.*, 1998 which has been adapted to the particular negotiation structures discussed in the previous section.



frame adjustment by	20 %	40 %
negotiation frame	8.8-17.2 ¢	10.2-15 ¢

frame adjustment by	20 %	40 %
negotiation frame	7.4-18.6 ¢	8.8-17.2 ¢

Figure 3. Exemplary Negotiation Frames and Adjustment

3.1. Reinforcement Learning: Reinforcement Learning is a computational approach for understanding and automating goal-directed learning and decision making. It focuses on individual learning from direct interaction with the individual's environment. Please recall from section 1 that Reinforcement Learning is different from supervised learning, the kind of learning studied in most forms of machine learning, statistical pattern recognition, and artificial neural networks. In contrast, for interactive problems it is often impossible to obtain examples of desired behavior that are both correct and representative of all the situations in which the agent has to act. In an unknown environment, where one would expect learning to be most beneficial, an agent must be able to learn *from its own experience only*.

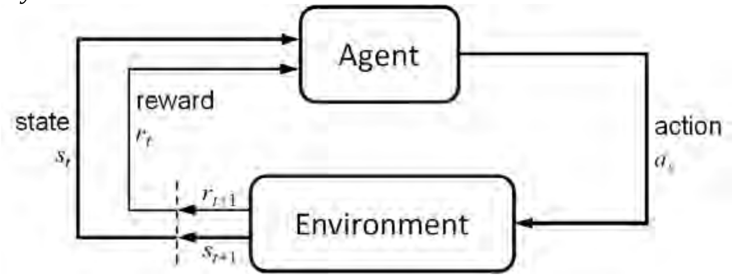


Figure 4. The Agent-Environment Interaction in Reinforcement Learning

In a Reinforcement Learning problem an agent (the learner and decision maker) interacts continuously with an environment that comprises everything outside the agent. Such an agent is faced repeatedly with a choice among different actions and corresponding responses by the environment which present new situations to the agent. After each action a *numerical reward* is received depending directly on the action taken. The agent's goal is to maximize the amount of accumulated reward in the long run (see figure 4, taken from Sutton *et al.*, 1998). More specifically, agents interact with the environment at discrete time steps. At every time step an agent evaluates some representation of the environment's *state* s_t. Based on that state it selects an *action* a_t. One time step later the agent receives a numerical *reward* r_t (probably a direct consequence of its action) and finds itself in a new state s_{t+1}. Based on this new situation the agent selects a new action a_{t+1} and so on and so forth.

One of the challenges that arise in Reinforcement Learning and not in other kinds of learning is the trade-off between *exploration* and *exploitation* when selecting new actions. To obtain a high reward, an agent will prefer actions that it has tried in the past, and found to be effective in producing reward. But for discovering favorable actions, it will equally explore actions that it has not selected before (in order to be prepared, or "educated", for changing situations to come). So the agent will *exploit* the past in order to obtain a reward, but it also will *explore* potential strategies in order to receive a high award in the future. The dilemma is that neither exploitation nor exploration can be pursued exclusively without failing at the task at hand (Sutton *et al.*, 1998). It is in a combination of both attitudes that an agent aims at performing "best" actions. In a stochastic way, each action will have to be tried many times to gain a reliable estimate of its expected reward. *Exploitation* has been found to be the right thing for maximizing the expected reward within the next period, but *exploration* may produce the greater total reward in the long run (Sutton *et al.*, 1998).

3.2. The Collaborative Learning Model in DECENT: Reinforcement Learning in DECENT has to be applied by each consumer and producer based on local information only, since no agent has global information. In DECENT the only information an agent is continuously faced with is its demand (or supply) and the negotiated energy price that is valid for the duration of next period (see section 2.1.7). The negotiation scheme developed in section 2.2 guarantees that best energy prices may only be negotiated within balancing groups on the lowest level, thus between regionally close customers (see formulas (7) through (9)). Hence, high or low energy prices, respectively, are a direct indication of contracts that have been closed on higher levels, due to inappropriate negotiation strategies (opening offers or bids and gradients of the negotiation curves, see formulas (1) through (4)). Thus, price feedbacks can be utilized by the agents to learn from and adequately adjust their strategies to changing supply situations in DECENT.

The negotiation functions defined in section 2.2 through equations (1) through (4), are characterized each by a tuple of the form (s_{1Ci}, $bid_{Ci}(0)$) or (t_{1Pj},$offer_{Pj}(0)$), for consumers C_i or producers P_j, respectively. The parameters urg_0, A_k, B_k as well as the shrinking factor are constants (see formulas (3) and (4) in section 2.2). The strategy parameters s_{1Ci} and t_{1Pj} are to be selected each from a finite sequence of equidistant values, so are $bid_{Ci}(0)$ *and* $offer_{Pj}(0)$. The intervals are determined so as to avoid *excessive bargaining* strategies under which contracts very likely will be settled on the highest level (reserve capacity) which is unfortunate for all parties involved. (We omit the technical details here. They are part of a Ph.D. thesis and will be subject to a separate ICT publication.) These 2-tuples uniquely define a *negotiation strategy* for C_i and P_j, respectively**.** Each strategy is meant to be operated for a period, and it will be revised it after each period.

Let the set of all feasible s_{1Ci} be $\boldsymbol{S}_{Ci}$, the set of all feasible t_{1Pj} be $\boldsymbol{T}_{Pj}$. Having in mind that $A_1 \leq bid_{Ci}(0) \leq A_1 + ½(A_1 + B_1)$ and $A_1 + ½(A_1 + B_1) \leq offer_{Pj}(0) \leq B_1$ let us denote the set of feasible bids and offers for C_i and P_j *by* $\boldsymbol{O}_{Ci}$ *and* $\boldsymbol{O}_{Pj}$, respectively. Then $\boldsymbol{S}_{Ci} \times \boldsymbol{O}_{Ci}$ and $\boldsymbol{T}_{Pj} \times \boldsymbol{O}_{Pj}$ are called the *strategy space for* C_i *and* P_j, respectively. Hence, the strategy space of an agent is the finite set of strategies, the agent chooses from at the beginning of each negotiation period.

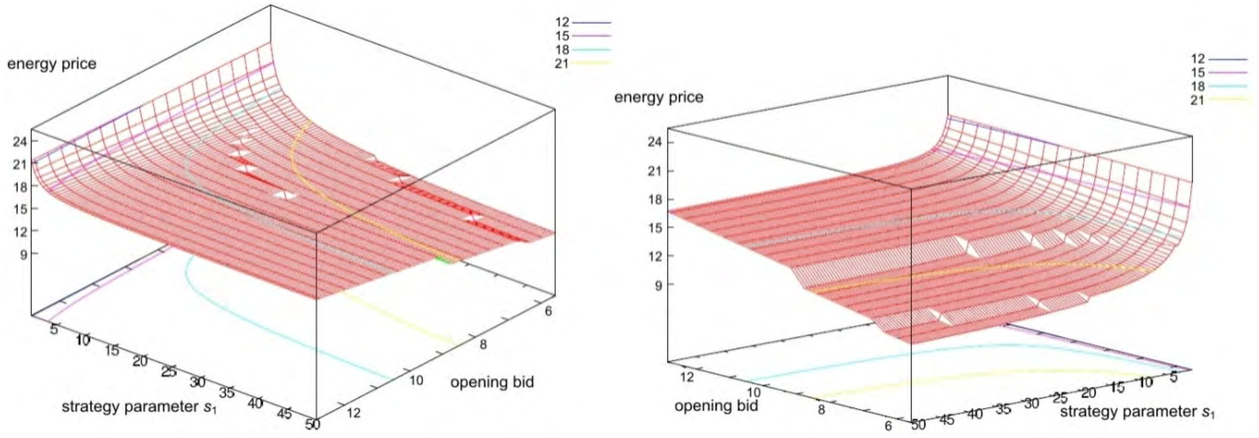


Figure 5. Varying Consumer Energy Prices (in ¢) throughout the Strategy Space

Figure 5 shows the consumer energy prices for varying strategies within a consumer's strategy space. The price relief depicted here was constructed during a balanced supply situation. In this scenario 25 producers and 24 consumers are negotiating with constant energy demands and static negotiation strategies (normally distributed opening bids and offers and strategy parameters s_1 and t_1, respectively). In this scenario a test consumer sequentially probes every strategy within its strategy space. Figure 5 depicts the resulting energy prices plotted against opening bids and values of parameter s_1.

However, in extensive experiments we have found that it is not reasonable for agents to regard the entire strategy space when exploring new and possibly better strategies. On the one hand the strategy space contains extreme strategies at its borders (extreme opening bids or offers with very steep or flat curves, respectively) that may only be successful if the system itself exhibits extreme demand or supply situations (e.g. critical undersupply or extreme surpluses of generated renewable energy). On the other hand it turned out that a multi-agent system of individual consumers and producers under distributed control and acting on local information alone does not arbitrarily changes its states. DECENT exhibits a certain inertia in its reaction to changes in supply situations: Successful strategies for unknown supply situations are likely to be "close" to the last known best strategy.

Taking the above phenomena into account we will introduce a neighborhood relation for the strategy spaces for C_i and P_j, respectively, as follows:

Two consumer strategies *strategy*$_{Ci}{}'$ and *strategy*$_{Ci}$ are *k-neighbors* iff (for a given parameter *k*):

$$\left|s'_{1C_i} - s''_{1C_i}\right| \le k \text{ and } \left|bid_{C_i}(0)' - bid_{C_i}(0)''\right| \le k \cdot \tag{10}$$

In the same way two producer strategies *strategy*$_{Pj}{}'$ and *strategy*$_{Pj}$ are *k-neighbors* iff:

$$\left|t'_{1P_j} - t''_{1P_j}\right| \le k \text{ and } \left|offer_{P_j}(0)' - offer_{P_j}(0)''\right| \le k \cdot \tag{11}$$

Hence, *k-neighborhoods* are designed as finite subsets of the strategy spaces $\boldsymbol{S}_{Ci} \times \boldsymbol{O}_{Ci}$ and $\boldsymbol{T}_{Pj} \times \boldsymbol{O}_{Pj}$, respectively. An energy price *r* achieved in a period is computed as the sum of the prices achieved for every quantum, including the price possibly paid, or received, at the reserve facility. In this way, after each period an agent is faced with a total energy price under the chosen strategy. Consumer agents try to keep their energy costs low while producer agents try to get reimbursed their investment and maintenance costs.

Strategic adaptations on behalf of C_i or P_j will be based on *normalized energy prices*. Formally speaking prices or negotiated within period *t* will be normalized according to:

$$r'_{C_i}(t) = \frac{r_{C_i}(t) - A_1}{B_1 - A_1} \text{ for consumers, and} \tag{12}$$

$$r'_{P_j}(t) = 1 - \frac{r_{P_j}(t) - A_1}{B_1 - A_1} \text{ for producers.} \tag{13}$$

3.2.1. Strategic Preference: In the DEZENT algorithm the individual consumption and production vary unpredictably (please remember from section 2.1 that it does not only depend on a customer agent whether or not he will act as a consumer, producer, or not act at all, during a period). Still customer agents want to keep their prices on a favorable level through the upcoming periods. This is pursued based on a judgment about the quality of strategy *a* selected for the current period *t*, in relation to the negotiated price *r*. For this purpose the *strategic preference for the next period* will be determined:

Definition: Let, for period *t,* and be the normalized prices negotiated under strategy *a*. The *strategic preference p*(*t*, *a*) is then defined for 2 different cases:

1) (regular strategic preference)

$$\begin{aligned} &p(0,a) := 0 \\ &p(t+1,a) := p(t,a) + \alpha(r'(t) - p(t,a)); 0 < \alpha \le 1 \end{aligned} \tag{14}$$

The step-size parameter determines how recent rewards are weighted against long-past rewards. In highly dynamic systems it makes sense to weight recent rewards more heavily than long-past ones, thus a higher value is deemed appropriate. In any case the step-size parameter will need to be adjusted specifically to the application at hand. Under certain circumstances is inappropriate to calculate the strategic preference as a weighted average over recent rewards. If the last known reward is far outdated it is not reasonable to calculate the strategic preference as the weighted average from those values and a most-recent reward (they may reflect completely opposing supply situations). Under such circumstances the strategic preference is set to the most recent reward:

2) (exceptional strategic preference)

$$\begin{aligned} &p(0,a) := 0 \\ &p(t+1,a) := r'(t) \end{aligned} \tag{15}$$

Over the periods, the strategies will be dynamically sorted by each agent according to their strategic preferences. During period t, the *currently best strategy* (to be selected for period t +1) is the strategy with the highest strategic preference value in the (dynamic) order.

There are 3 different modes for selecting the strategy for period t +1: *Exploitation*, *Explore₁* and *Explore₂*:

1) *Exploitation* selects the currently best strategy.
2) *Explore*$_1$ randomly picks a strategy which is in the k-neighborhood of the currently best strategy.
3) *Explore*$_2$ randomly picks any strategy.

3.3. The Collaborative Learning Algorithm in DECENT (DECOLEARN): Every customer agent executes the following algorithm:

1) Initially the strategies will be chosen randomly according to *Explore*$_2$ (or may be prescribed by the customers).
2) For selecting the strategy for period t +1; $t \geq 1$ the selection modes are given the following probabilities: $P(Explore_1) = \varepsilon_1$, $P(Explore_2) = \varepsilon_2$, $P(Exploitation) = 1\text{-}(\varepsilon_1 + \varepsilon_2)$ with $0 \leq \varepsilon_1, \varepsilon_2 \leq 1$ and $0 \leq 1\text{-}(\varepsilon_1 + \varepsilon_2) \leq 1$. According to these probabilities *exactly one* of the modes is randomly determined.
3) Through the determined mode, the strategy for period t +1 is selected, and will be executed.
4) If the mode was *Exploitation* or *Explore*$_1$ then the *regular strategic preference* for the next period will be computed (see formula (14)). In case of *Explore*$_2$ the *exceptional strategic preference* will be used (see formula (15)).

For an example we refer the reader to our simulation studies (see table 1).

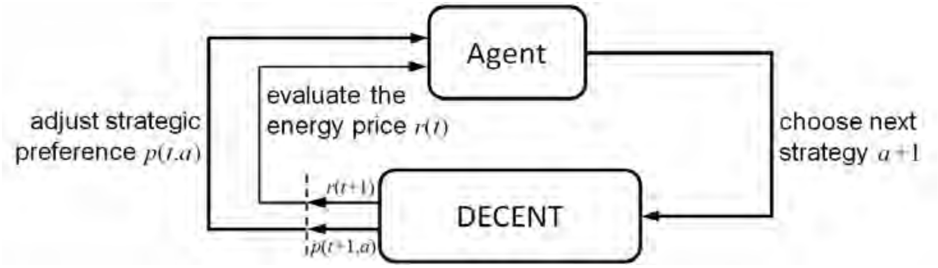

Agent
evaluate the energy price $r(t)$
adjust strategic preference $p(t,a)$
choose next strategy $a+1$
$r(t+1)$
DECENT
$p(t+1,a)$

Figure 6. The Agent-Environment (DECENT) Interaction in DECOLEARN

Figure 6 depicts the adjusted Reinforcement Learning model in DECOLEARN (please compare it to figure 4). The only way for an agent in DECENT to interact with its environment and to generate a reward feedback is by bidding on energy quantities and evaluating the negotiated energy price. The reward is used to adjust the strategic preference of the appropriate action. Based on the adjusted preferences a new strategy for the following period is chosen.

Figures 7.a through 7.d depict the dynamic adjustment of preferences for “best” strategies in a producer’s strategy space. In this example the situation shifts from a balanced system-wide demand and supply to a shortage of energy generation on the lower negotiation levels (see figures 1 and 3). The producer starts from a normally distributed sampling of strategies within his strategy space (in a balanced demand and supply situation, see figure 7.a). During the undersupply producers learn to favor negotiation strategies with high opening offers and high strategy parameters t_1 (flat negotiation curves, see figures 7.b through 7.d).

Although the producer depicted in figure 7 favors strategies with highest opening offers, he does not at the same time choose extreme values for his strategy parameter t_1. The producer learns to omit this combination of highest opening offers in combination with extremely flat negotiation curves. Even in undersupply situations, with these negotiation strategies he risks getting propagated onto higher negotiation levels and ultimately satisfies his needs at the backup reserve facility, thus at lowest possible rates (see figure 3). However, a producer may still choose these extreme negotiation strategies randomly in the course of an *Exploration*$_2$ action choice.

4. Simulation Studies

In order to evaluate the impact of distributed learning on negotiations in DECENT we compared the DECOLEARN algorithm with the previously introduced algorithm for which the strategies remained unchanged across the periods. We term it here the *STATIC* algorithm.

In previous publications (Wedde *et al.*, 2008) it has been demonstrated that a very large majority of customers got settled on level 1 (see figures 1 and 3). Into the same vein it has been argued that early satisfaction within a period was beneficial for all parties involved. In particular, the flexibility for selecting negotiation parameters was better: The intervals (A_k, B_k) are larger for smaller k (see section 2.2) thus the exponential curves for producers and consumers would converge faster yet at less favorable rates, for both of them. Our comparative studies were consequently based on 2 major criteria:

1) *Minimize, if not avoid, the access to the main reserve capacity* (see figure 1);
2) Allow the customers for contracting as early within a period as possible.

We conducted a series of experiments with 3000 agents and a layered agent negotiation structure exhibiting 3 levels (for the experimental setup see table 1). The experiments were conducted on the high-performance *LiDO* cluster at the Technical University of Dortmund, over 7200 periods (=1 hour) with realistic profiles for individual producers and consumers of renewable energy. The resulting demand and supply pattern utilized in our experiments is depicted in figure 8. Every energy quantum not successfully negotiated on these levels is sold to, or purchased at, a system-wide reserve facility.

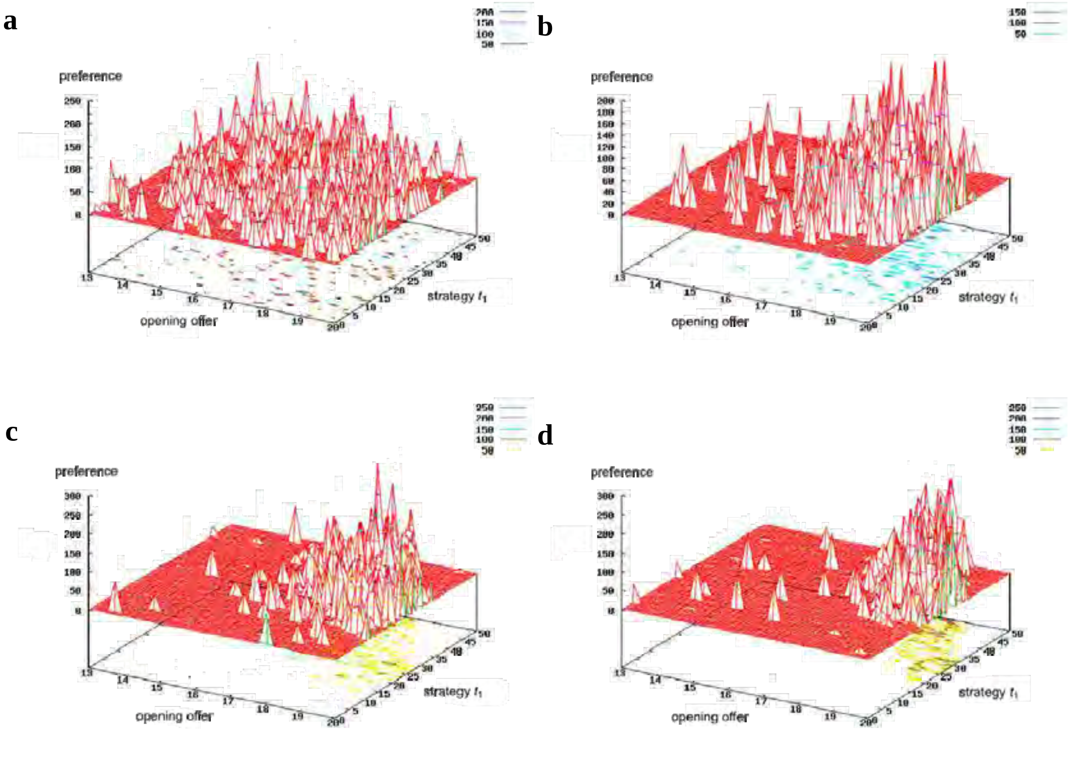


Figure 7. Development of Strategic Preferences Values

Producers are modeled based on real world profiles derived from wind turbines and photovoltaic installations of various capacities (50-300 W/h). Consumers are modeled based on empirical studies on average power requirements of households. (A detailed description of the tools utilized in modeling realistic behavior of producers and consumers of renewable energy is beyond die scope of this presentation yet this is subject of a separate publication).

Table 1. Experimental Setup

Negotiation Levels	3
Number of BGMs on Level 1	30
Number of BGMs on Level 2	12
Number of BGMs on Level 3	1
Number of Clients per BGM on Level 0	100
Number of Consumers (50 W/h average capacity)	2160
Number of Producers (50-300 W/h average capacity)	840
Simulation Time (negotiation periods)	7200
$A_1 = 5$ ¢, $B_1 = 20$ ¢, *similarity* = 1.2 ¢, *allowance* = 50 W, *shrinking rate* = 20 %	
energy refund = 4 ¢/W, *energy price* = 21 ¢/W (reserve facility)	
$\underline{s}_{1C_i}, \underline{t}_{1P_i} = 0.5$, $\overline{s}_{1C_i}, \overline{t}_{1P_i} = 50$	
$\alpha = 0.3$, $k = 2$, $\varepsilon_1 = 0.45$, $\varepsilon_2 = 0.05$	

The demand and supply pattern shows an overall energy demand that fluctuates around 120 kW/h with an amplitude of approximately 10 kW/h (see figure 8). The overall power production oscillates smoothly with a period of approximately 2000 seconds. In the course of the simulation the overall production rises from 70 kW/h to 140 kW/h. Thus, for the first 4300 periods the overall energy demand slightly exceeds the overall energy production. After 4300 periods the overall energy demand and supply situation improves as it features a production surplus throughout the remainder simulation time. A third curve at the bottom of figure 8 resembles this imbalance in overall demand and supply.

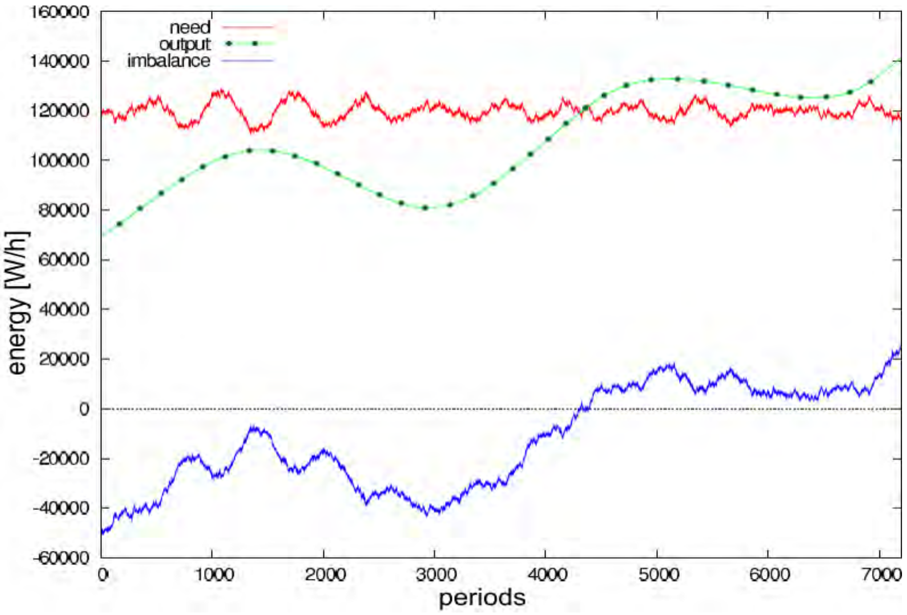


Figure 8. Overall Demand and Supply Pattern

In order to evaluate the performance of distributed learning in DECENT, the overall renewable energy contracted on the 3 negotiation layers before reaching the system-wide reserve facility is plotted against the simulation time (see figure 9). The results are compared against the maximum amount of energy that can be purchased under ideal negotiations (reflecting the global power balance before imbalances are left with the reserve facility as indicated in figure 3). This "ideal" curve coincides with the minimum of both the overall production and the overall demand (see figure 8). Imbalances in demand and supply beyond level 3 have to be regulated at the system-wide backup capacity (level 4, at 380 kV in figure 3).

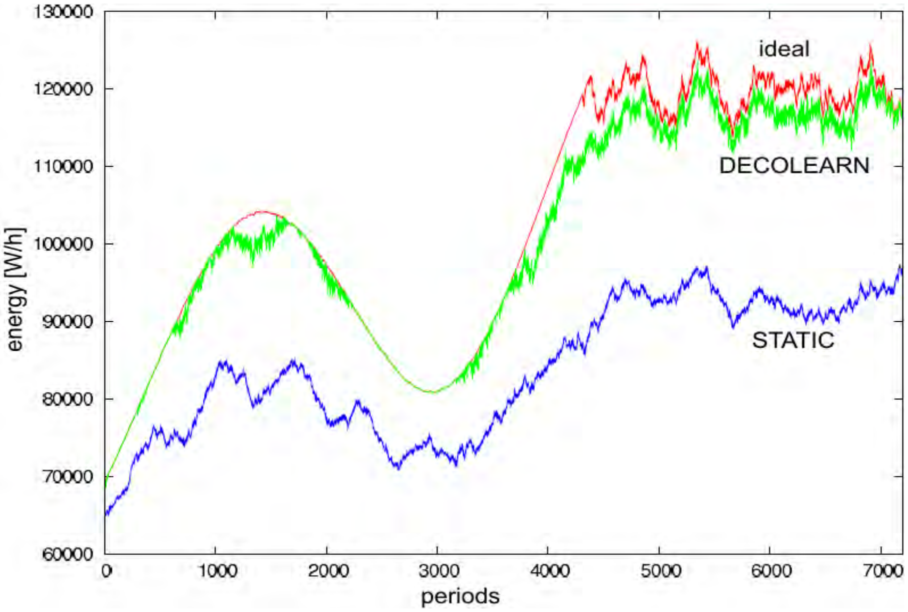


Figure 9. Energy Purchased throughout the Multi-Level Negotiations in DECENT

In the course of these experiments the DECOLEARN algorithm has been evaluated against the distributed STATIC algorithm. For this purpose additional experiments have been conducted on the basis of the setup in Table 1. Instead of dynamic strategy adjustments after each period, strategies have been chosen at random once at the beginning of the experiment, and then kept constant throughout the 7200 periods. Simulations have been repeated 50 times with different initial strategy settings. The resulting averaged curve is depicted in figure 9.

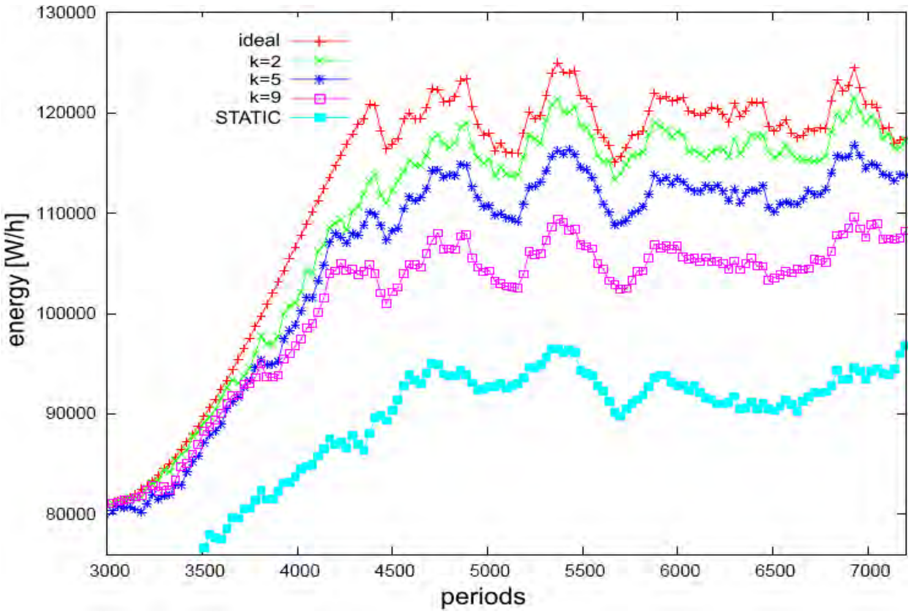


Figure 10. Experiments with Various Configurations of *k*

With the chosen setup the negotiations based on static strategies are unable to adjust to varying supply situations and thus perform significantly worse than ideally possible throughout all 7200 periods. This is illustrated in the gap between the STATIC curve and the ideal curve that the algorithm is unable to close. This “underperformance” increases after approx. 4000 periods when overall system dynamics increase. Distributed learning performs nearly ideally, and far better than STATIC, throughout the entire simulation. After approximately 4000 periods the performance of DECOLEARN changes. The maximum amount of energy that can be sold before reaching the backup facility fluctuates more rapidly as it has to adapt to the high variation of the demand curve (see figure 8). So the performance of DECOLEARN decreases somewhat (although it remains still far superior to STATIC). A reasonable interpretation may come from problems in finding “correct” or “good” adjustments of *strategic preferences* within the agents’ strategy neighborhoods.

In order to indicate the influence of the size of parameter *k* (neighborhood width, see section 3.2) on the DECOLEARN performance we repeated experiments with different values for *k*. The results between the periods 3000 and 7200 are depicted in figure 10.

With growing *k* DECOLEARN exhibits problems while coping with the increasing fluctuation in the maximum amount of energy that can be sold before reaching the backup facility. This supports the conjecture made in section 3.2 that distributed learning in large neighborhoods is suffering from an increased uncertainty about “good” adjustments under more rapidly varying supply and demand situations. It is not reasonable for agents to regard the entire strategy space when searching for better strategies. Successful strategies for unknown supply situations are likely to be within a narrow *k*-neighborhood of the last known best strategy.

Regarding the criterion of *early satisfaction on the lowest negotiation level* we assessed the impact of distributed learning on the amount of energy negotiated on the lowest level. Figure 11 depicts the amount of energy sold and purchased on the first level for both DECOLEARN and STATIC strategies. The positive DECOLEARN performance trend that was already observed in figure 9 results to a large extent from its stronger performance on the first negotiation level, throughout the whole simulation. We found that negotiations on the second and third level performed alike for DECOLEARN and STATIC yet with a relevant advantage for DECOLEARN, after 4000 periods.

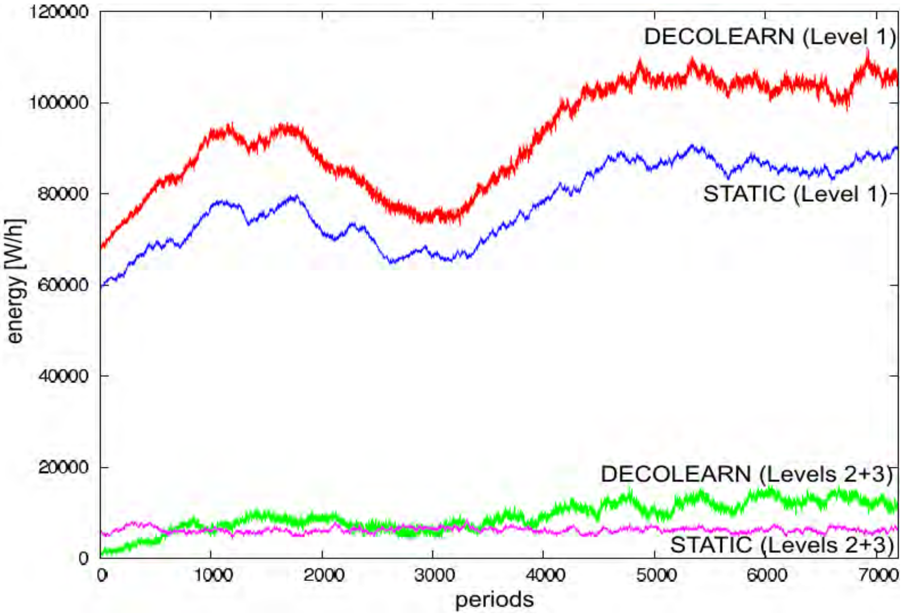


Figure 11. Negotiated Energy on the Levels 1, 2 and 3

5. Conclusion

The general idea behind DECOLEARN as derived from Reinforcement Learning was that individually favorable prices should be adaptively found, and at the same time such rates should be dynamically achieved during later periods as well.

Due to page limitations, we restricted ourselves to a representative report on our experimental evaluation. However, due to the distributed control concept, the results do not depend on the number of agents involved: Even with 300.000 agents (simulated on the HPC Cluster at the Technical University of Dortmund) we found the picture to be equally favorable.

In the model scenario there is a large amount of uncertainty on different levels, last but not least due the fact that for every agent, be it a producer or consumer, it is unpredictable whether it will act as a producer or consumer agent, or remain inactive during the next period. So at first glance the STATIC algorithm may well have its own merits as a very simple framework for handling unpredictable negotiations dynamically. Compared to $Explore_1$ or $Explore_2$ as degenerative forms of DECOLEARN, i.e. for $P(Explore_1)=\varepsilon_1=1$, $P(Explore_2)=\varepsilon_2=0=P(Exploitation)$, or $P(Explore_2)=\varepsilon_2=1$, $P(Explore_1)=\varepsilon_1=0=P(Exploitation)$, STATIC may be expected to exhibit a similar behavior, depending on the amount of variation or customer unpredictability. In this way STATIC served as a natural and simple candidate for the (comparative) evaluations.

We have argued above that a bottom-up management is most adequate for the widely dispersed sources of electric power. It should also be very clear from the previous discussion that a distributed algorithmic solution and the corresponding computer network support are indispensable for the purpose.

The results show very clearly the strong improvement that comes into the picture through the flexible and adaptive distributed learning methods in DECOLEARN. With DECOLEARN we successfully introduced a highly dynamic concept for negotiations across periods and, indeed, for providing the macro-customers (human or technical actors) with decent solutions at reasonable costs (see figure 3 and Wedde *et al.*, 2008), despite the unpredictability of all actions and influences.

In the DECENT initiative we have pursued a novel distributed approach for negotiating, distributing, and monitoring electric energy in large power grids. While in the world of liberalized power markets traditional power management concepts have come to their limits, considering both in achieving optimal prices, in avoiding power line overloads, and in maintaining grid stability (absence of regional or global black-outs), the rapidly emerging regenerative power production – while offering an ecologically highly desirable future – poses even stronger challenges on a traditional grid management, in all aspects discussed above. In this situation we have defined a distributed control approach which combines, and completely integrates, measures and solutions in the economic, ecologic, electric, and ICT aspects. Distributed agents take care of producer and consumer needs in a bottom-up fashion, balancing at the same time the needs within balancing groups and beyond. Stability is taken care of through novel distributed algorithms Krause *et al.*, 2009, allowing for testing and achieving a stable distribution pattern *for each period*. Peak demand and supply management have been successfully handled in a distributed procedure, an extension of the DEZENT negotiation algorithm Wedde *et al.*, 2008. Please recall that the DECENT bottom-up approach allows for a high amount of parallelism which is the basis for the timely micro-operations of the agents while allowing getting along without look-ahead strategies, spot market procedures etc.

Our experimental results so far reach to the 110 kV grid level, however, conceptually there are no limits to expand into larger dimensions and grids interconnected by even higher voltage levels.

References

Ahamed T.P.I., Rao P.S.N. and Sastry P.S. 2002. A Reinforcement Learning Approach to Automatic Generation Control. *Electric Power Systems Research*, Vol.63, No.1, pp.9-26.

Amin S.M. and Wollenberg B.F. 2005. Toward a Smart Grid: Power Delivery for the 21st Century. *IEEE Power and Energy Magazine*, Vol.3, No.5, pp.34-41.

Anthony P., Hall W., Dang V. and Jennings N. 2001. Autonomous Agents for Participating in Multiple Online Auctions. *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI'01) Workshop on E-Business and the Intelligent Web*, Vol.1, No.1, pp.54-64.

Bakirtzis A.G. and Tellidou A.C. 2006. Agent-Based Simulation of Power Markets under Uniform and Pay-as-Bid Pricing Rules using Reinforcement Learning. *Proceedings of the Power Systems Conference and Exposition (PSCE '06)*, Vol.1, No.1, pp.1168-1173.

Burnett M.B. and Borle L.J. 2006. A Power System Combining Batteries and Super-Capacitors in a Solar/Hydrogen Hybrid Electric Vehicle. *Proceedings of the 2006 IEEE Vehicle Power and Propulsion Conference*, Vol.1, No.1, pp.709-715.

Ernst D. and Wehenkel L. 2002. FACTS Devices Controlled by Means of Reinforcement Learning Algorithms. *Proceedings of the 14th Power Systems Computation Conference (PSCC 2002)*, Vol.1, No.1, pp.1-7.

Ernst D., Glavic M. and Wehenkel L. 2004. Power Systems Stability Control: Reinforcement Learning Framework. *IEEE Transactions on Power Systems*, Vol.19, No.1, pp.427-435.

Greenwald A.and Kephart J.O. 1999. Shopbots and Pricebots. *Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI '99)*, Vol.1, No.1, pp.506-511.

Hadidi R. and Jeyasurya B. 2009. Near Optimal Control Policy for Controlling Power System Stabilizers Using Reinforcement Learning. *Proceedings of the 2009 IEEE Power & Energy Society General Meeting*, Vol.1, No.1, pp.1-7.

Hommelberg M.P.F., Warmer C.J., Kamphuis I.G., Kok J.K. and Schaeffer G.J. 2007. Distributed Control Concepts Using Multi-Agent Technology and Automatic Markets: An indispensable feature of smart power grids. *Proceedings of the 2007 IEEE Power Engineering Society General Meeting*, Vol.1, No.1, pp.1-7.

Krause O., Lehnhoff S., Rehtanz C., Handschin E. and Wedde H.F. 2009. On Feasibility Boundaries of Electrical Power Grids in Steady State. *International Journal of Electric Power & Energy Systems*, Vol.31, No.9, pp.437-444.

Nanduri V. and Das T.K. 2007. A Reinforcement Learning Model to Assess Market Power under Auction-based Energy Pricing. *IEEE Transactions on Power Systems*, Vol.22, No.1, pp.85-95.

Potter C.W., Archambault A. and Westrick K. 2009. Building a Smarter Smart Grid through Better Renewable Energy Information. *Proceedings of the IEEE Power Systems Conference and Exposition*, Vol.1, No.1, pp.1-5.

Pipattanasomporn M., Feroze H. and Rahman S. 2009. Multi-Agent Systems in a Distributed Smart Grid: Design and Implementation. *Proceedings of the IEEE Power Systems Conference and Exposition*, Vol.1, No.1, pp.1-8.

Sutton R.S. and Barto A.G. 1998. *Reinforcement Learning*. MIT Press.

Takano T., Matsumoto K., Oki, I. and Ohashi T. 1996. A Case Learning Tool for Operations of Power Systems. *Proceedings of the*

International Conference on Intelligent Systems Applications to Power Systems, Vol.1, No.1, pp.91-96.

Tellidou A.C. and Bakirtzis A.G. 2006. Multi-Agent Reinforcement Learning for Strategic Bidding in Power Markets. *Proceedings of the 3rd International IEEE Conference on Intelligent Systems*, Vol.1, No.1, pp.408-413.

Vlachogiannis G. and Hatziargyriou N.D. 2004. Reinforcement Learning for Reactive Power Control. *IEEE Transactions on Power Systems*, Vol.19, No.3, pp.1317-1325.

Wedde H.F., Lehnhoff S., Rehtanz C. and Krause O. 2008. Bottom-Up Self-Organization of Unpredictable Demand and Supply under Decentralized Power Management. *Proceedings of the 2nd IEEE International Conference on Self-Adaptation and Self-Organization*, Vol.X, No.X, pp.74-83.

Wedde H.F., Lehnhoff S., Moritz K.M, Handschin E. and Krause O. 2008. Distributed Learning Strategies for Collaborative Agents in Adaptive Decentralized Power Systems. *Proceedings of the 15th IEEE International Conference on Engineering of Computer-Based Systems*, Vol.1, No.1, pp.26-35.

Ye H., Brocco A., Kuonen P., Courant M. and Hirsbrunner B. 2008. SmartGRID: A Fully Decentralized Grid Scheduling Framework Supported by Swarm Intelligence. *Proceedings of the 7th International IEEE Conference on Grid and Cooperative Computing*, Vol.1, No.1, pp.160-168.

Zhou M., Ren J., Li G. and Xu X. 2003. A Multi-Agent based Dispatching Ooperation Instructing System in Electric Power Systems. *Proceedings of the 2003 IEEE Power Engineering Society General Meeting*, Vol.1, No.1, pp.436-440.

Biographical notes

Prof. H. F. Wedde received his diploma degree in Pure Mathematics, the Ph.D. degree in Computer Science (1974), all from the University of Bonn, Germany. His appointments included a Senior Research Staff position at the Gesellschaft für Mathematik und Datenverarbeitung (GMD) in Bonn, Germany (1969-83), visiting professor positions at the universities of Pisa (1980) and Turin (1982), Italy, and at the Academy of Sciences in Warsaw, Poland (1977, 1980). Prof. Wedde was on faculty of Wayne State University/ Detroit (1984 - 1993). From 1994 on, he has been with the School of Computer Science at the University of Dortmund, Germany. His major research interests are in various areas of Distributed Computing Systems including Theory, Distributed Operating/ File Systems, Distributed Real-Time Systems, Distributed Security, and Nature-Inspired Routing algorithms, the latter especially in the Bee Hive project. Recent efforts have been devoted to such application areas as distributed (secure and real-time) power management of renewable energy, integrated production and transportation scheduling in Distributed Manufacturing, or traffic jam avoidance.

Dr. S. Lehnhoff received his diploma degree in Computer Science in 2005 and his Ph.D. in 2009 from the Technical University of Dortmund, Germany. He is a research associate with the Institute of Operating Systems and Computer Architecture at Technical University of Dortmund, Germany. His research interests include but are not limited to smart grids, intelligent power management systems and distributed control algorithms for applications in collaborative self-organizing systems.

Prof. C. Rehtanz received his diploma degree in Electrical Engineering in 1994 and his Ph.D. in 1997 at the University of Dortmund, Germany. From 2000 he was with ABB Corporate Research, Switzerland and from 2003 Head of Technology for the global ABB business area Power Systems. From 2005 he was Director of ABB Corporate Research in China. From 2007 he is professor and head of chair for power systems and power economics at the Technical University of Dortmund. His research activities include technologies for network enhancement and congestion relief like stability assessment, wide-area monitoring, protection, and coordinated FACTS- and HVDC-control.

Dr. O. Krause received his diploma degree in Electrical Engineering in 2005 and his Ph.D. in 2009 at the Technical University of Dortmund, Germany. He is a research associate at the Institute of Electrical Power Systems and Power Economics, Technical University of Dortmund, Germany. His research interests are strategies for the coordinated use of power systems and issues related to network stability.

Received December 2009
Accepted March 2010
Final acceptance in revised form March 2010