Power Load Management as a Computational Market

Fredrik Ygge
Department of Computer Science (IDE)
University of Karlskrona/Ronneby
372 25 Ronneby, Sweden
Fredrik.Ygge@ide.hk-r.se, http://www.pt.hk-r.se/~fredriky

Hans Akkermans
AKMC & University of Twente
Department of Information Systems (INF/IS)
P.O. Box 217, NL-7500AE Enschede, The Netherlands
akkermans@ecn.nl, J.M.Akkermans@cs.utwente.nl

Abstract

Power load management enables energy utilities to reduce peak loads and thereby save money. Due to the large number of different loads, power load management is a complicated optimization problem. We present a new decentralized approach to this problem by modeling direct load management as a computational market. Our simulation results demonstrate that our approach is very efficient with a superlinear rate of convergence to equilibrium and an excellent scalability, requiring few iterations even when the number of agents is in the order of one thousand. A framework for analysis of this and similar problems is given which shows how nonlinear optimization and numerical mathematics can be exploited to characterize, compare, and tailor problem-solving strategies in market-oriented programming.

1 Introduction

Computational markets have been suggested as a solution to resource allocation, scheduling, and optimization problems (e.g. [2][5][7][8]). In this paper we present how computational markets can be used to handle the application of power load management. A load, in this context, is any device that consumes electric energy, such as a water heater or an electric motor. With load management we mean the concept of controlling the loads at the demand side to achieve a better use of energy, better for the utility, the customer or both. Load management can be used for many different purposes, like avoiding overloads in bottlenecks in the grid, reduce losses caused by reactive power, reduce overtones and stabilize the network. One normally distinguishes between two different categories of load management [4]: direct and indirect. Direct load management implies that the utility determines what loads are to be connected, reduced, or disconnected at specific occasions. Indirect load management is the case where the utility sends some signal to the customer, such as price information, and relies on his/her ability to adjust to this signal. The topics discussed in this pa-
is outlined in section 4, and conclusions are given in section 5.

2 HOMEBOTS: Agents for Automated Load Management

From the energy utility point of view it is desirable to have a system that hides most details of the different loads while still providing enough information for energy optimization. The system should be able respond to high level control commands for, e.g., reduction of the current load in the distribution system by a certain amount.

2.1 Computational Market Design

Every controllable load in our system is represented by a load agent, called a HOMEBOT, whose needs and preferences are modeled by a utility function, describing how important it is for the agent to have a share of the resource (i.e. the electrical power). From these utility functions we introduce market mechanisms which settle the distribution of energy according to the market model described below. The agents are grouped in correspondence with the inherent topology of electrical grid and the communication system. This can for example be as in Figure 1.

![Diagram](image)

Figure 1 Agent society structure

The role of the interface agents, as shown in Figure 1, is to provide the utility with an interface to the load management system. The interface agents can be thought of as the utility’s local salesmen. An interface agent can thus be directly manipulated by the utility in order to make the system behave in a certain way, so they embody the decentralized control strategies of the energy utility. Interface agents might or might not at the same time serve as load agents. Two examples of the use of the interface agents are given in section 3.

Note that even though we have introduced a computational market, the relation between the utility and the customers is regulated through contracts, and the purpose of the computational market is to make good use of these contracts.

2.2 Utility Functions

The utility function is determined by a number of factors, e.g. load model, current state of the load, user model, utility/customer contract and expected future prices.

From our experience it is, for a number of loads and contracts, very normal to have a concave utility function, so the marginal utility (i.e. the first derivative of the utility function), which serves as the bidding price, decreases with increased resource (i.e. the second derivative is negative). This property of the utility function is very common in other application areas as well [7][13].

Figure 2 shows two typical utility functions, one representing a water heater and one representing an electrically heated house. In the case of the water heater the utility increases until the heater is fully heated. For a heated house, on the other hand, the utility typically raises to a certain point, when comfort is maximized for the user, and then gradually decreases.

In this paper we have chosen a utility function, denoted by \( u(r) \), where \( r \) is the allocated resource (electric power) for a certain agent, of the form\(^1\):

\[
(1) \quad u(r) = a - b \cdot e^{-(c \cdot r)} - (d \cdot r)
\]

By adjusting the parameters, \( a, b, c, \) and \( d \), we can describe a large number of useful utility functions. For each load, \( r \) is constrained, in that it has a lower and an upper bound.

---

1. This utility function is chosen as a reasonable utility function. The exact form will be established from further investigations of load models etc.
2.3 Bids and Auction Schemes

Reallocations in computational markets are often realized through auctions (e.g. [2][5][7]) and this mechanism is also adopted here. In an auction every agent in a group sends a bid, to a designated auctioneer or to the other agents, and then a reallocation of the resource is computed. When designing a computational market, one has to decide how bids should be constructed and, from these bids, determine a proper reallocation scheme. Thus, the auctioneer should maximize the total utility of the system from the current bids of the agents. Formally this leads to the following optimization problem:

\[
\text{(2) Maximize } \sum_{j=1}^{n} u_j(r_j) \\
\text{such that } \sum_{k=1}^{n} r_k = R
\]

Since the total utility is the sum of the local utilities for every agent, the problem is said to be separable [6]. This property gives rise to a significant simplification of the reallocation task.

From nonlinear optimization theory we know that when the Kuhn-Tucker conditions are fulfilled, i.e. when all first derivatives of the utility functions, \(u'_j(r_i)\), are equal to a shared value \(\lambda\), for all agents having an amount of resource different from their boundary values, then we have achieved an optimal global solution for this constrained nonlinear separable concave problem. (For details refer to for example Ibaraki and Katoh [6] or Fletcher [1].) Loosely, this means that no agent which is not at its boundary, has a higher marginal utility than any other agent.

It follows that an optimal solution can be obtained by solving the nonlinear equation system \(u'_j(r_i) - u'_n(r_n) = 0\), for \(i = 1, 2, ..., n-1\). To this end, we introduce a column vector \(f = [f_1, f_2, ..., f_{n-1}]^T\), where \(f_i = u'_j(r_i) - u'_n(r_n)\). Then we apply the standard multi-variable Newton method, (3), for solving the vector equation \(f = 0\):

\[
\text{(3) } \Delta r^k = (\text{grad } f^k)^{-1} f^k \\
r^{k+1} = r^k + \alpha \Delta r^k
\]

where \(r\) is the column vector, \(r = [r_1, r_2, ..., r_{n-1}]^T\), grad \(f\) is the gradient matrix of \(f\) (also called Hessian matrix, containing the second-order partial derivatives of the utility function), \(\alpha\) is a step size, and \(k\) is the index of the iteration. With utility functions of the above kind we can use a step size equal to unity. The key advantage of a Newton scheme is its very fast quadratic convergence; a property clearly borne out in our experiments, and of central importance to real-life load management.

For solving the constrained problem we rewrite \(r_n\) as \(R - r_1 - r_2 - ... - r_{n-1}\). Then grad \(f\) is:

\[
\text{(4) } \text{grad } f_j = \begin{cases} u''_j(r_j) + u''_n(r_n) & i=j \\ u''_n(r_n) & i \neq j \end{cases}
\]

After the elimination-by-substitution of the resource constraint, we have an unconstrained optimization left, which we solve by the standard Newton scheme cf. equation (3). Due to the separability, inversion of the Hessian matrix in equation (3) is easily done in closed form by use of the Sherman-Morrison formula [3].

In order to manage the local boundaries of the resource we use a scheme based on the RELAX algorithm with lower and upper bounds on variables [6] with the above Newton method as relaxed allocation algorithm. Loosely, the idea of the RELAX algorithm with lower and upper bounds on variables is to first solve the relaxed allocation problem, i.e. without bounds on variables, and for each of the variables that exceed their limit value assign the corresponding limit value and allocate the remaining resource among the unlimited agents by using the RELAX algorithm recursively. For details refer to Ibaraki and Katoh [6]. It is not difficult to show that this way of dealing with the local bounds theoretically scales at most linearly with the number of agents. (In our simulation experiments of section 3, it even appears to be rather independent of the number of agents.)

The information needed from every agent in each auction, the agents' bids, is the current value and the lower and upper bounds of the resource and the first and second derivatives of the utility function. Then for every consecutive iteration of the auction the first and second derivatives are needed.

With the simple utility functions specifically used in equation (4) we could also use an alternative price directed auction scheme, see section 4. However, the Newton algorithm will work for a very wide class of utility functions, and is thus more general.

3 Simulation Results

In this section we present simulation results demonstrating how interface agents can be used to control the loads in the distribution system, and see that the above algorithm rapidly
reallocate the resource properly in the system. Throughout the simulation, one iteration is performed per time unit. As we want to take advantage of the benefits of really decentralized control, the energy utility behaves like any other party on a competitive market, rather than wielding monopolistic power as in a purely centralized approach. We will see that this indeed turns out to be possible, and very efficiently so. Looking at the formal problem statement of equation (2), it follows that basically two different strategies are open to the utility within the confines of distributedness: 1) Influencing a single interface agent’s resource in the global resource constraint, so as to effectively reduce the total R. 2) Influencing the utility optimum, by changing a single interface agent’s utility function and, thereby, bidding price.

Simulation results for both strategies are described in sections 3.1 and 3.2 respectively. For reasons of clarity, results are given for a system of ten agents. Scaling issues for the field application are discussed in section 3.3.

3.1 Reducing by a Certain Amount of Resource
In the first simulation we demonstrate a situation where the utility needs to reduce the total load, R, in the system, due to shortage in production, overload in a bottleneck in the grid, or similar. In this case the interface agent is designated to a load, say a water heater with a minimum load of 0kW and maximum load of 3kW. The utility can directly remove or add resource to this agent. Here we demonstrate how a request for a reduction of 5.3kW is managed by the system.

In Figure 3 we see that from a given initial allocation, measured in kW of electric power, the system rapidly relaxes to equilibrium (in two time units). Then at time six the utility requests the reduction and simply sets the going resource value for the interface agent from approximately 1.3kW to -4kW. This results in a non-feasible allocation, since the lower limit of the interface agent (0kW) is violated. First, the system regains a feasible allocation; after one time unit the utility controlled load is above its lower bound (0kW), due to reduced allocation to other agents, and thus the 5.3kW (1.3 + 4) reduction is performed. After another three time units the system is back in equilibrium. If the price of the resource directly corresponds to the real price of power, we immediately can tell how much this load reduction costs. This is of course very useful when evaluating the load management system.

3.2 Changing the Price
As a second control strategy we show a similar system, but one where the interface agent is not a load agent. Here the utility attempts to influence the total energy consumption not by controlling a specific resource, but by bidding a specific price through changing its utility function. It then trusts the load management system to achieve a new equilibrium; here it is open how much the customer loads will be reduced. For achieving this we have used a linear utility function for the interface agent. This means that the interface agent will “buy back” power, thus reducing the consumption of the loads and decreasing production with the same amount, from all agents having a lower marginal utility than the marginal utility of the interface agent, i.e. the slope, or derivative, of the straight line.

In Figure 4 a simulation of a price change is shown. Again the agents are assigned a given initial amounts of resource. The interface agent (the top curve in Figure 4) starts with a marginal utility (its bid price) of say 0.3. From the initial situation, equilibrium is reached in three time units. Then the utility attempts to reduce the consumption of the loads, by entering a new bid price for its interface agent equal to, say.
just a matter of seconds. The more so because the tackled application problem. As the number of successful applications quickly increases, there is a growing need for a common framework for analysis, as the ones shown above. We note that in a real-life application, the number of loads will typically be in the order of a few hundreds in a secondary substation area. For getting back to equilibrium when the disturbance is introduced, three to four and two iterations are needed in simulation one and two respectively. The result is much more favorable than the worst-case theoretical result which indicates a linear scaling with the number of agents. In the simulations we have only referred to time in terms of number of iterations, but in field applications we have to deal with real time constraints. However, a realistic value for reaching a new equilibrium, which depends on communication speed, computational power in the network, and agent configuration, is estimated to be below 0.5s. This is an extremely satisfactory figure, because it means that our market approach to power load management in real-life situations is just a matter of seconds.

4 Analysis

Above, we have shown the practical value of the computational market in solving a specific, but important, real-life application problem. As the number of successful applications quickly increases, there is a growing need for a common framework for analysis, the more so because the tackled problems as well as the problem-solving strategies appear highly diverse. Application-independent, knowledge-level characterizations of problems and strategies are needed, in order to gain an improved understanding of, and formulate better guidelines for market-oriented problem-solving. For computational markets, the following knowledge level aspects of analysis seem particularly useful to us:

1) Problem or overall task characteristics: goals, givens, working conditions and features set by the task environment.
2) Task organization: decomposition, I/O flow, control, responsibilities due to the distribution over agents.
3) Information requirements: what is needed to know for successful task execution, and where is this information located?
4) Communication requirements: what information must be communicated to other agents, and when?
5) Agent competencies: what must be/is an agent capable of?
6) Performance and quality metrics for task execution. This enables us to 'position' different approaches and strategies in a multidimensional 'space' of analysis.

Wellman [15] emphasizes a similar point, and specifically proposes to build on concepts and theory of microeconomics as a formal foundation, and Kurose and Simha [7] offer a practical example. These are steps in turning the computational market idea from an attractive but informal metaphor for multi-agent system architectures (the famous 'invisible hand'[11]) into a full scientific theory. We concur with such a position, but note that there is an extensive additional body of knowledge in modern mathematics, which in our view has not yet been exploited to the full benefit of formal analysis of market-oriented programming. This concerns in particular nonlinear optimization and numerical mathematics (for recent overviews, see e.g. [1][6]). In this section we sketch some insights ensuing from such an analysis, illustrated by various computational market problems published in the literature.

4.1 Resource- vs. Price-Oriented Strategies

Our task organization - which we see as the technical layout of what in fact is a new business process in energy utilities - is based on what is sometimes called a resource-directed strategy, in contrast to a price-directed strategy. Essentially, this distinction refers to how different subtask responsibilities are distributed between the trading agents on the one hand, and the auctioneer on the other hand. The dynamic variables to be dealt with in market problems are given by the vector pair (p,r) of bid prices and demands for resources. In a price-directed strategy, the trading agents handle the r-part by indicating their resource demands at the going price (by inversion of the first derivative of the utility function), and the auctioneer then deals with the p-part, repeatedly sending out market update information in terms of adjusted price signals until the resource constraints are met. This leads to a market dynamics whereby the prices function as the independent variables. In resource-directed strategies, it is the other way around. Here the resource equality demand is first dealt with (through elimination by substitution), and then the utility optimization task is undertaken (by an iteration procedure).

It can be verified that the computational complexity in both cases is the same under appropriate conditions. A capa-

4.2 Bid Information and Convergence

As pointed out in section 2.3 we have made the strategic choice that load agents communicate two information point items \((p, p')\) consisting of their marginal utility at the current local resource and its derivative. In principle, one item is sufficient for a working solution, but adding the second item has the advantage that it leads to a quadratic rather than linear convergence rate. If this competency of obtaining derivative values is not available, in our application we could switch to variable-metric (quasi-Newton) or conjugate-gradient strategies which will need more iterations but are still superlinear.

We note that most current market schemes, including the binary search in Walras as well as ‘tatonnement’ in general [14], the mixed scheme for energy management in [5], and the first-order algorithm of [7], are all linear schemes, because they are all in some way proportional to the current ‘error’. Communication requirements are also very different: Walras [14] communicates full curves rather than point items, the information content in [5] is different but the related communication burden is higher (since additional global information is required for the local bids) compared to our case, whereas in the first-order algorithm of [7] only a single point item is communicated. It follows that given the amount and nature of the communicated information, in all mentioned systems except the first-order algorithm of [7], superlinear rather than linear convergence speed could be achieved. More generally, many strategic choices are at our disposal that can improve problem-solving and computational properties. The tools of nonlinear optimization and numerical mathematics are helpful in a formal analysis of this, and we will show this by example in 4.3 and 4.4 in some technical detail.

Another point worth mentioning here is that in Walras, since the entire utility function is sent to the auctioneer, the convergence is a task solely resting with the auctioneer, while the other strategies presented here require inter-agent communication at each iteration.

4.3 Analysing Kurose-Simha

Kurose and Simha [7] consider a distributed file allocation problem. The major difference with our problem is in the absence of upper bounds and in the shape of the utility functions. In Kurose and Simha’s application it is taken to be

\[
u_i (r_i) = c_i r_i + \frac{r_i}{\mu_i - \lambda_i r_i}
\]

where \(c_i, \mu_i, \lambda_i\) are cost parameters

Their simulation results demonstrate clearly how computational properties favorably change by adding a single communicated information item \((p', p)\) in going from their first-order to their second-order algorithm. These results become theoretically obvious by our following proposition.

Proposition: The second-derivative algorithm of [7] is equivalent to a Newton algorithm (as outlined in our equation (3)), specialized to an optimization problem that (i) is separable and (ii) has a linear equality constraint.

Proof: Separability leads to a simple diagonal Hessian matrix in unconstrained problems, and to the form of our equation (4) in the constrained case. Due to this special form, equation (3) can be analytically solved by using the matrix inversion formula named after Sherman-Morrison [3]. Carrying out these calculations precisely gives Kurose and Simha’s second-derivative algorithm.

Corollary: Being a Newton scheme, this is provably a quadratically convergent algorithm, in the sense that quadratic test functions are optimized in \(O(1)\) iterations. In contrast, linear schemes as mentioned above, need \(\Omega(n)\) iterations for solving the Kuhn-Tucker gradient conditions, \(n\) representing the number of dimensions, i.e. agents, computer nodes or commodities. The utility functions of both equations (1) and (5) are sufficiently similar to a quadratic function over a wide range so that a Newton scheme converges and practically achieves this superlinear convergence. This also theoretically explains why our HOMEBOTS system attains equilibrium in so few iterations.

Thus, a proper choice of information and communication resources and competencies leads to very efficient problem-solving. This has been shown above within a re-
source-directed strategy, but also holds for price-directed strategies. Interestingly, Kurose and Simha emphasize the drawbacks of the price-directed strategy. However, we now prove that their file allocation problem might also be efficiently solved by a price-directed strategy which is equivalent to Walras [14].

**Proposition:** The (unconstrained) resource demands corresponding to the (first derivative of the) utility function (5) are given by

$$r_i(p) = \frac{\mu_i}{\lambda_i} - \frac{1}{\lambda_i} \sqrt{\frac{\mu_i}{p - c_i}}$$

**Proof:** The corresponding Kuhn-Tucker gradient condition (all dimensions/agents share the same market equilibrium price \( p \) for their marginal utilities) results, due to the separability, in a set of equations with a single resource variable and \( p \) as a free parameter. These can be easily solved by function inversion. Basically, the resource scales with the inverse square root of the market price, as seen in equation (6).

**Corollary:** The task of the auctioneer is to find a value for the equilibrium price \( p \) such that the resources satisfy the global equality constraint (telling that (6) sum up to unity). This is a one dimensional root search problem. We further note that our application might also be handled in the same fashion. Analytical results, such as these, are very useful for testing and validating.

### 4.4 Analysing Walras

Along the same lines we can show that Wellman's Walras system [14][15] might solve the same problems through alternative resource-directed strategies. In his price-based strategy he employs binary search for the global resource constraint, which has a linear rate of convergence. This may be improved to a quadratic scheme, especially if second-order derivative calculations can be made. In his application to configuration design [15], the constant elasticity utility function is used, and in this case this is not much of a problem.

Walras has the special feature that it solves a multi-commodity problem, while all other cited work as well as our own application involves a single commodity. This poses the question [14] how the complexity of resource allocation problems scales with both the number of agents and commodities. A partial answer was given in the previous sections, where we pointed out that worst-case theory gives a linear scaling with the number of agents, for concave separable problems, and for each commodity. In practice, we found that it is even more favorable. This fits, by the way, with Wellman's remarks. He conjectures that the scaling with the number of commodities is much more badly. Nevertheless, we have that mutatis mutandis, interchanging agents with commodities, a worst-case statement holds analogous to the one above. In Walras, each agent has to solve an optimization problem, but this is very efficiently done by inverse function evaluation. For each commodity, we have a global resource constraint. In the Walras formulation [15], however, these are all coupled through the prices. This leads to a root search in \( k \) dimensions simultaneously, with \( k \) denoting the number of commodities, rather than \( k \) times a one dimensional search as in a separable case. Hence, the Walras approach is separable in terms of agents, but nonseparable in terms of commodities. This explains the difference in computational complexity. But other strategies with different scaling properties are conceivable.

### 5 Conclusions

In this paper we introduced a computational market design for power load management. Power load management aims at managing the energy consumption of customer equipment such as boilers and heaters. Every device acts as an agent in a computational market buying and selling energy. This is a decentralized way to reduce unwanted peak loads and hereby save money. We describe an operational simulation environment, called the HOMEBOTS system. Simulation experiments carried out with the system show that:

- the optimal allocation is achieved within a few bidding iterations.
- the system scales up nicely since the number of needed iterations is rather independent of the number of agents, even if it is as high as one thousand.
- the utility can act as being just a party among many on the market, and still rely on very simple tactics to successfully manage many loads: it suffices to either remove a certain amount of resource or to change the bidding price, impersonated by a single utility-controlled agent.

In conclusion, market-oriented programming is well suited for this kind of applications, because the information required for optimization is inherently distributed, problem-solving turns out to be simple and efficient and it enables loads to be added and deleted without modifications of the existing system.

Furthermore, we have given a brief, broader analysis of this and other computational market problems from the literature. Nonlinear optimization and numerical mathematics
provide an integrated framework to characterize different market strategies and to compare their advantages and disadvantages. From our analysis we conclude that:

- this branch of mathematics offers solutions to problems much more complicated than currently investigated in market-oriented programming, opening up new application areas.
- it is possible to find schemes with a superlinear rate of convergence for market-oriented programming problems, better than the often used binary search schemes.
- it is possible to derive analytical insights into computational markets, important for validating and testing computational market programs for real settings.

Hence this approach helps to better understand, but moreover, considerably extends available problem-solving strategies in market-oriented programming.

Acknowledgments

We would like to thank professor Rune Gustavsson at the University of Karlskrona Ronneby and president Hans Ossossen at EnerSearch AB for their large support and encouragement. Hans Olsson at Lund University for enlightening us on some numerical analysis issues, and the SoC team at the University of Karlskrona/Ronneby and Lund University, the load management team at Lund University, the student team Flux at SIKT for interesting discussions and comments on draft material.

References