Best-Response Planning of Thermostatically Controlled Loads under Power Constraints*

Frits de Nijs, Mathijs M. de Weerdt, and Matthijs T. J. Spaan

Delft University of Technology
The Netherlands
{f.denijs, m.m.deweerdt, m.t.j.spaan}@tudelft.nl

Abstract
Renewable power sources such as wind and solar are inflexible in their energy production, which requires demand to rapidly follow supply to maintain an energy balance. Promising controllable demands are heat buffers that use electricity to maintain a temperature at a setpoint. Such Thermostatically Controlled Loads (TCLs) have been shown to be able to follow a power curve using reactive control. In this paper, we investigate the use of planning under uncertainty to pro-actively control an aggregation of TCLs to overcome temporary imbalances. We present a formal definition of the planning problem under consideration, which we model using the Multi-Agent Markov Decision Process (MMDP) framework.

Since we consider hundreds of agents, solving these MMDPs directly is intractable. Instead, we propose decomposing the problem by decoupling the interactions through arbitrage. Decomposition of the problem means relaxing the joint power consumption constraint, which means that joining the plans together can cause overconsumption. Arbitrage acts as a conflict resolution mechanism during policy execution, using the future expected value of policies to determine which TCLs should receive the available energy. We experimentally compare several methods to plan with arbitrage, and conclude that a best response-like mechanism is a scalable approach that returns near-optimal solutions.

1 Motivation
A promising potential storage capacity can be found in the various heat buffers operated by consumers: houses, refrigerators and hot water reservoirs all need to be maintained at a certain temperature offset from the environment temperature, to which they decay over time. This requires constant action to counteract. The exact moment when these buffers are heated or cooled, however, can often be shifted. By storing more energy in the heat buffer now, we can later ‘extract’ this heat from the buffer in the form of reduced loads. In this sense, we may think of heat buffers controlled by thermostats as a kind of batteries (Hao, Sanandaji, Poolla, & Vincent, 2013).

*The full paper has been accepted for publication in Proceedings of the Conference on AAAI (de Nijs, Spaan, & de Weerdt, 2015).
These Thermostatically Controlled Loads (TCLs) can themselves be controlled automatically by the system operator. Existing work has mainly focused on using Control Theory to make an aggregation of TCLs closely follow an available power signal for balancing on a timescale of seconds (Callaway, 2009; Hao et al., 2013).

Instead of using TCLs for balancing, we investigate the potential of TCLs for buffering temporary drops in power production, such as those caused by the fluctuations in renewable generation. These fluctuations typically occur on longer timescales (Koch, Zima, & Andersson, 2009), potentially violating the consumers’ comfort constraints if no preventive action is taken. To respond to fluctuations before they occur, we propose a planning approach with an objective to minimize the discomfort experienced by end-users. In our Power Constrained Planning (PCP) problem formulation, the deviation from the TCL setpoint is optimized under power availability constraints. Because power constraints may be too severe to guarantee comfort, we aim to minimize the total discomfort.

We show empirically that solving optimal formulations of this problem directly is intractable. Therefore we present a decomposition of the Multi-agent Markov Decision Process (MMDP) model (Boutilier, 1996). We decouple the problem and let each agent compute its own individual plan using a single-agent MDP. Combining the individual plans will result in a joint plan that might not be feasible due to power constraints. Hence, we introduce an arbitrage mechanism that ensures power constraints are respected by limiting the number of TCLs that can be switched on simultaneously. However, by in turn computing best-response policies and by learning the probability of being allowed to switch on, we compute scalable and effective solutions to the PCP problem. Using a real-world inspired scenario we show that the best-response policies are able to scale to realistic problem sizes, while maintaining a solution quality that is near optimal.

2 Power Constrained Planning

A thermostatic load is any device that is able to consume (electric) power for the heating or cooling of a body in relation to the outdoor temperature, such as refrigerators or central heating systems. The goal of the thermostat is to operate the device such that the temperature of the body remains as close as possible to a given setpoint at all times.

A Markov chain model of thermostats controlled by hysteresis controllers was presented by Mortensen and Haggerty (1988). In their model, the temperature of the body in the next time step $\theta_{i,t+1}$ is deduced from the current temperature $\theta_{i,t}$, the current outside temperature $\theta_{\text{out},t}$, and the binary on/off decision $m_{i,t}$. Temperature function $f(\theta_{i,t}, m_{i,t}, \theta_{\text{out},t})$ models inertia by preserving a fraction $\alpha$ of the current temperature (depending on the device specific insulation level). The remaining fraction $1 - \alpha$ is used to trend towards either outside temperature when off, or some maximum temperature depending on the heating capacity.

We adapt this unconstrained model into a Power Constrained Planning (PCP) problem as follows. Given a known, time-variable limit $L_t$ on the sum of devices switched on, we set out to solve the problem of determining an activation schedule for a collection of TCLs. We use boldface characters to represent vectors of device parameters over all devices. Then, given a planning horizon $h$ and the device parameters, we set out to find a device activation schedule that minimizes the discomfort cost function $c(\theta_t)$. The
entire planning problem becomes:

\[
\begin{align*}
\text{minimize} & \quad \sum_{t=0}^{h-1} c(\theta_t) \\
\text{subject to} & \quad \theta_{t+1} = f(\theta_t, m_t, \theta_{\text{out}}) \\
& \quad \sum_{i=1}^{n} m_{i,t} \leq L_t \\
& \quad m_{i,t} \in [0, 1] \quad \forall i, t
\end{align*}
\]

Due to the generality of the model, the controlled loads can have different objectives which can be expressed through the cost function. As a representative example of a cost function we use a variant of the squared error where minor (0.5 degrees) offsets incur no costs at all:

\[
c(\theta_t) = \sum_{i=1}^{n} \max\{0, |\theta_{i,t} - \theta_{\text{set}}| - 0.5\}^2.
\]

We use this cost function because it has been shown in user studies by de Dear and Brager (1998) that 90% generally acceptable comfort levels can be found in a small (2.5°C) band surrounding the ideal temperature.

This model can be straightforwardly encoded using either a Mixed-Integer Programming (MIP) or a Multi-agent Markov Decision Process (MMDP) approach. However, both approaches are expected to suffer exponential scaling in runtime with the number of agents. Therefore, we propose an approximation based on decoupling the power constraint in the next subsection.

### 2.1 Decoupling, Arbitrage, and Best-Response

The only interaction among the TCLs is the allocation of the power availability. We therefore introduce a method that exploits this limited level of interaction as follows. First, we decouple the agents and let them find an individually optimal plan. Then we use an arbitrage mechanism upon execution to make sure that the agents together do not overuse the available power. Finally, we show how to let the agents coordinate their plans by simulating the arbitrage mechanism in advance, and performing a best-response to each other’s policies. Each of these steps is discussed in more detail below.

First we completely decouple the agents, i.e., we factorize the MMDP by discarding the restriction \(L_t\) in the joint transition function. Then we solve a single-agent MDP for each agent separately.

Second, to overcome the problem that the agents’ plans are no longer guaranteed to jointly stay within the power limit, we resolve the conflicts at policy execution using an arbitrage mechanism: in case too many devices want to switch on in a certain time step \(t\), we iteratively search for the agent that expects to lose the least utility from switching off, and switch it off. This is repeated until the conflict is resolved. To determine which agent expects to lose the least utility by going from on to off we look at the difference between the planned utility scores in the value table.

Third, the plans for the single-agent MDPs take into account the effect of arbitrage and thereby indirectly of the plans of other agents by estimating how likely an agent is to be assigned such a constrained action.
The first step towards estimating the probability of being assigned energy is to take the pessimistic assumption that everyone always wants to make use of all available energy. Given this assumption, the probability of being able to activate the TCL is \( p_{on}(t) = \frac{n_t}{n} \). This allows agents to detect potentially congested regions in time, but without knowing the actual level of congestion. We call this method Pessimistic in the experiments.

To estimate the actual demand in congested regions, we can simulate policies computed in an earlier iteration and tally how much energy is requested. Then agents can learn their relative probability of receiving power, by counting the frequency with which his request for energy is granted in a time step \( t \). Its probability then becomes \( p_{on}(i, t) = \frac{\text{received}_{i,t}}{\text{requested}_{i,t}} \). This method is labeled as Adaptive in the experiments.

### 3 Experimental Results

To empirically evaluate the scalability of the proposed decomposition versus optimal solving techniques we compare the runtime performance of both approaches on small generated instances. Since performance is expected to depend on the number of agents and the length of the horizon, we control for these variables. We generated instances with the horizon fixed at 20 and the number of agents increasing from 1 to 6, and with the horizon from 5 to 45 and the number of agents fixed at 3 (10 instances per setting). To keep times manageable, a run-time cut-off of 5 minutes is imposed. The average run-time performance is shown in Figure 1.

As expected, the run-times of the optimal methods scale exponentially with the number of agents. On top of this, the MIP runtimes scale exponentially with the length of the horizon, because this approach is not able to fully exploit the Markov property of the temperature progression. The proposed decompositions do not appear to exhibit this exponential scaling. Therefore we expect that the adaptive approach is scalable to the size of typical real-world instances.

To demonstrate this, and to evaluate the resulting solution quality, we apply the adaptive approach to a simulated neighborhood of 182 households equipped with heat-pumps. The simulated parameters are modeled to match measured performance observed by van Lumig (2012, page 60). To determine if the solution quality is near opti-
For this instance, we compare the pessimistic and the adaptive decomposition to a relaxation of the optimal MIP. This relaxation allows the devices to be switched on partially. Figure 2 presents the average indoor temperature, the normalized cumulative error and the number of devices switched on for this instance.

The cumulative penalty of the adaptive decomposition stays close to the MIP relaxation lower bound, which confirms our expectation that the adaptive decomposition performs close to optimal. In addition we see that the pessimistic decomposition heats up too much. This is evident from the cumulative penalty because the amount of error incurred above the deadband is higher than below it. Nevertheless, both planning solutions perform much better than the non-anticipatory control. Finally, it is important to note that computing the 182 policies for the adaptive decomposition took only 7.5 minutes, less time than it took the optimal MMDP solver to compute a solution for some five agent instances. This demonstrates that the adaptive decomposition is indeed scalable to real-world instances.

4 Conclusions and Future Work

In this paper we investigate planning the activation of Thermostatically Controlled Loads (TCLs) for buffering for periods of low power availability. To this end we present a planning problem definition for the optimal control of Thermostatically Controlled Loads under power constraints. Because optimal solution methods do not scale on this problem definition, we propose a decoupling of the Multi-agent Markov Decision Process (MMDP) into an MDP per agent. At policy execution time, any conflicts that occur are resolved using an arbitrage mechanism, which greedily uses the policies to determine which agents benefit most from the available energy. By using a best-response planning
process, agents are able to learn how likely they are to receive energy in a certain time step. From evaluating the adaptive decomposition on both large and small instances, we conclude that it is able to return near-optimal solutions while remaining scalable. This method therefore seems promising to control an aggregation of TCLs.

To extend this work we are considering two directions. On the one hand, we currently assume that the power production curve is known. In practice we do not know exactly how much power will be available for the TCLs in the future. Thus, we want to investigate planning TCLs with uncertain power production curve. An advantage of the MDP framework is that it naturally allows for encoding uncertainty through the transition function, which makes the existing approach a promising starting point. On the other hand, we want to look into the addition of power quality constraints. Switching on many TCLs on a single feeder cable causes a large voltage drop for the last TCL in line. This can cause violations of the power quality norms, and risks damaging electric appliances. Hence, planning should be extended to take into account not just power but also voltage constraints in the grid.

5 Acknowledgments

Support of this research by network company Alliander is gratefully acknowledged.

References