Front. Eng. Manag. 2017, 4(3): 304–314
DOI 10.15302/J-FEM-2017045

RESEARCH ARTICLE

Zhaohui ZHANG, Ruilong DENG, Tao YUAN, S. Joe QIN

Sliding window games for cooperative building temperature control using a distributed learning method

Abstract In practice, an energy consumer often consists of a set of residential or commercial buildings, with individual units that are expected to cooperate to achieve overall optimization under modern electricity operations, such as time-of-use price. Global utility is decomposed to the payoff of each player, and each game is played over a prediction horizon through the design of a series of sliding window games by treating each building as a player. During the games, a distributed learning algorithm based on game theory is proposed such that each building learns to play a part of the global optimum through state transition. The proposed scheme is applied to a case study of three buildings to demonstrate its effectiveness.

Keywords game theory, demand response, HVAC control, multi-building system

1 Introduction

Buildings, which consume about 70% of the total electricity generated in the US, are among the largest energy consumers in the power grid (Weng and Agarwal, 2012). According to data of energy consumption in European households (Dounis and Caraiscos, 2009), 68% of the energy consumption of buildings comes from space heating or cooling, 14% from water heating or cooling, and 13% from electric appliances and lighting. Aside from accumulated energy use, buildings tend to have a high electricity demand simultaneously, thereby causing significant peak demand exertion on the grid (Ma et al., 2011; Ma et al., 2012). Therefore, electricity price usually varies positively with the peak demand to curtail potential overload and grid instability during peak load periods. One of the main goals of advanced building control systems under the big picture of demand-side management (Deng et al., 2015; Mohsenian-Rad et al., 2010; Li et al., 2011; Deng et al., 2014; Chai et al., 2014; Zhang et al., 2016) is the minimization of the overall energy cost from space heating or cooling in response to electricity prices rather than simply minimizing energy consumption.

Thermal comfort in a working or living place is mandatory to ensure the satisfaction and productivity of the occupants. The improvement of building comfort usually demands high energy consumption. Hence, one of the most important issues for smart and energy-efficient buildings is to ensure thermal comfort while minimizing energy cost or reducing peak-hour energy usage (Levermore, 2000).

The supervisory control level in the literature of smart building heating, ventilation, and air conditioning control systems usually aims to reduce energy cost while maintaining the desired indoor comfort level (Wang and Ma, 2008; Deng et al., 2016). Extensive research efforts have been made with regard to energy-efficient building comfort management. Various approaches can be roughly classified into two categories, namely, conventional and computational intelligence methods (Dounis and Caraiscos, 2009; Shaikh et al., 2014). Conventional methods include proportional-integrate-derivative (PID) control, which solves overshoots in thermostats with a dead zone (Levermore, 2000); optimal control, which maintains control performance while further reducing energy cost (Zaheer-Uddin and Zheng, 2000; Kummert et al., 2001); adaptive control, which enables self-regulation and

Received May 30, 2017; accepted August 12, 2017

Zhaohui ZHANG, Tao YUAN
Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089, USA

Ruilong DENG
Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 1H9, Canada

S. Joe QIN (✉)
Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089, USA; The Chinese University of Hong Kong, Shenzhen 518172, China
E-mail: sqin@usc.edu

adaptation to climate conditions (McCartney and Nicol, 2002); and model predictive control (MPC), which is proposed to introduce prediction horizons and models for future disturbances (Oldewurtel et al., 2010; Široký et al., 2011; Ma et al., 2014). Computational intelligence methods include neural network approaches, fuzzy logic schemes, and evolutionary algorithms, which usually include user participation in the specification of the desired comfort. Moon and Kim (2010) designed a thermal control logic framework with four thermal control logics, including two predictive and adaptive logics using neural network models. Dounis and Manolakis (2001) proposed general guidelines for the design of a fuzzy logic thermal comfort regulator when multiple input/output controlled systems exist. Kolokotsa et al. (2002) and Wright et al. (2002) proposed genetic algorithms, which are derivative-free and require minimal specific information.

However, most of the work in the literature focuses on characterizing and optimizing a single building and uses centralized methods (Levermore, 2000; Deng et al., 2016; Kummert et al., 2001; Ma et al., 2014; Moon and Kim, 2010; Dounis and Manolakis, 2001). Inspired by distributed control and multi-agent research in other areas (Dimarogonas et al., 2012; Fan et al., 2013; Li et al., 2013), cooperative energy management for multiple smart buildings from a distributed perspective is investigated while considering their preferences and leveraging their flexibilities. From the perspective of a consumer, the coordination of all buildings is desirable to cooperatively achieve global benefit, thereby resulting in overall global optimum while satisfying the need of each building.

The global benefit of building energy cooperative management has three aspects (Zhang et al., 2014; Zhang et al., 2017). First, buildings are anticipated to cooperate with one another such that the total load of all buildings remains below a certain threshold. The threshold is related to the distribution infrastructure capacity, such as the capacity of transformers and feeders within the set. When the capacity is exceeded, a penalty will be added for compensation and adjustment. Second, the minimization of global energy cost, which is a function of real-time electricity price and total power load, is desired. The price of electricity that varies with time can be exploited to reduce the cost of consumption. Third, the required comfort level of indoor temperature for each building, that is, the room temperature, should not fall outside the comfort zone. In this paper, a weighted average of the aforementioned three aspects of benefit is formulated as the global utility function.

The problem in the maximization of global utility function can be solved using convex optimization methods in a centralized way, with a central controller being used to handle all the buildings. However, solving the problem in a distributed manner, which has several advantages over centralized control, is of great significance. First, a distributed control structure ensures that systems are reliable. In a distributed system, the performance of other buildings is virtually unaffected when a controller breaks down for one building. By contrast, all buildings will be severely affected if a centralized controller breaks down. Second, a distributed control system is scalable. The system can be constructed at a large scale and scattered in a large area. It provides convenient infrastructure when a new building is built and becomes part of the existing control system. Furthermore, the optimization and computation load for each controller would be significantly reduced, thus allowing distributed computing and storage.

In this paper, the distributed global utility maximization problem is formulated as a series of sliding window games, in which each building is considered as a player, the global utility is decomposed to the payoff of each player, and each game is played over a control horizon. In each game, a newly proposed distributed learning algorithm based on game theory (Marden et al., 2014), which teaches each player to play part of the Pareto optimum by state transition, is applied. During the games, each player maximizes its own payoff based on the action it played and the payoff it received without knowing others. Overall, the distributed algorithm can achieve a global near-optimal solution, that is, the solution converges to the centralized optimal solution with a probability approaching one.

The rest of this paper is organized as follows. Section 2 describes the centralized model. Section 3 reformulates the multi-building decision making problem by using a sliding window and game theory payoff functions, and the distributed learning algorithm is applied. Section 4 provides case studies for an hourly game and a horizon of hour game to compare the performance. Section 5 draws the conclusions.

2 Centralized model

A group of n buildings denoted by the set $N = \{1,2,...,n\}$ and a centralized controller is considered to perform an overall optimum control behavior.

The global economic objective function is

$$\phi(L,P) = \sum_{i=1}^{n} \sum_{t=1}^{24} c_i\Big(l_i(t),\ p(t)\Big), \tag{1}$$

where $\phi(.)$ is the aggregate daily electricity expense of all buildings, $L = [L_1,...,L_n]^{\mathrm{T}}$ is the energy consumption matrix for n buildings, and $L_i = [l_i(1),...,l_i(t),...,l_i(24)]^{\mathrm{T}}$ is the energy consumption column vector for building i, where $l_i(t)$ is the power consumption of building i at tth time step. $P = [p(1),...,p(24)]^{\mathrm{T}}$, where $p(t)$ is the hourly electricity price per energy unit. The function $c_i\Big(l_i(t),p(t)\Big) = c_1 p(t) l_i(t)^2 + c_2 p(t) l_i(t) + c_3$ models the operation cost of building i with power load $l_i(t)$ and price $p(t)$.

At every time step, the temperature of each building should stay within a comfortable range and the total power load should stay below a threshold

$$T_{i\_lb} \leqslant T_i(t) \leqslant T_{i\_ub}, \forall i \in \{1,...,n\}, t \in \{1,...,24\},$$
$$\sum_{i=1}^{n} l_i(t) \leqslant L_r, \forall t \in \{1,...,24\}.$$

The functional relationship of the next time step indoor temperature prediction with respect to the current time step indoor temperature, outdoor temperature, and power consumption can be described according to Forouzandehmehr et al. (2013).

$$T_i(t+1) = \varepsilon T_i(t) + \Big(1-\varepsilon\Big)\Big(T_{\mathrm{OD}}(t) - \gamma K l_i(t)\Big), \tag{2}$$

where ε is the thermal time constant of the building, γ is a factor that captures the efficiency of the air conditioning unit, K is a conversion factor, and T_{OD} is the outdoor temperature. The indoor and outdoor temperatures are usually known. Thus, a corresponding expected room temperature for the next hour $T_i(t+1)$ is observed once an energy consumption amount is determined.

Therefore, the centralized model is as follows:

$$\begin{aligned} &\min_{L} \sum_{i=1}^{n}\sum_{t=1}^{24} c_i\Big(l_i(t),\ p(t)\Big) \\ &s.t. \sum_{i=1}^{n} l_i(t) \leqslant L_r,\ \forall t \in \{1,...,24\} \\ &T_{i\_lb} \leqslant T_i(t) \leqslant T_{i\_ub},\ \forall i \in \{1,...,n\},\ t \in \{1,...,24\}. \end{aligned} \tag{3}$$

In the centralized decision making problem, our objective is to minimize the total daily electricity cost of the group of buildings. Decision variables are the amount of energy consumption $l_i(t)$ for every building i at every hour t or equivalently, the temperature set-point for each building in every hour. Each building should satisfy two constraints: the total power load in every hour should not exceed a certain threshold, and the temperature should be within a comfort region.

The centralized decision making problem can be solved using convex optimization techniques. In the problem, $p(t)$, L_r, and $T_{\mathrm{OD}}(t)$ are pre-known parameters. The objective function is the sum of a set of quadratic functions whose coefficients are usually positive. Therefore, this function is a convex objective function. The total power load constraints are linear. If the temperature comfort level constraints are transferred from the functions of $T_i(t)$ to the functions of $l_i(1), l_i(2), ..., l_i(t)$, then all constraints are linear. Therefore, many convex optimization solution algorithms, such as interior point method, can be used.

However, the problem is aimed to be reformulated into a multi-building decision making problem because of the drawback of centralized control in comparison with distributed control. That is, each building can behave as an autonomous agent while interacting within the group. Moreover, each building should learn to play a part of the optimal solution without having any information on the system and on the operations of other buildings. This condition leads us to deterministic game theory.

3 Game reformulation and distributed cooperative control

In this paper, the multi-building decision making problem is reformulated as a game, with each building being treated as a player, and a newly proposed payoff-based distributed learning algorithm (Marden et al., 2014) in game theory is applied to solve this problem. The solution is a near-optimal solution or the solution converges to the optimal solution with a probability that is infinitely close to one.

To construct a sliding window game, the sliding window energy consumption vector for building i at the tth time step $y_i(t)$ can be defined as

$$y_i(t) = [l_i(t)...l_i(t+N_p-1)]^{\mathrm{T}}, \tag{4}$$

where $l_i(t)$ is the power consumption of building i at tth time step, N_p is the width of a fixed-width sliding window, and (N_p-1) is the length of the prediction horizon.

At each time step, although an optimal power consumption strategy for the whole N_p time intervals within the sliding window time frame is computed, only the solution of the current time interval should be adopted to the output because even though the prediction of future inside temperature based on Eq. (2) is valid for the next time step, it will not be as accurate when making a series of predictions. When predicting the inside temperature of the $(t+1)$th time step, we use the updated real-time inside and outside temperature information of the tth time step instead of the predicted inside temperature of the tth time step based on the $(t-1)$th time step and the predicted outside temperature.

The optimization procedure will be repeated and similar games will be formulated in subsequent time intervals as the window slides. The constraints are formed as penalty functions and slight changes are applied in our game formation in the next section because the game theory tool cannot directly solve the constrained problem. Another point is for long time frame and multiple buildings, the decision space for the distributed learning algorithm is too large. Thus, we first formulate the problem as hourly game and then extend it to multiple hour sliding window game. In the following section, the reformulation of the problem is described from the game theory perspective.

3.1 Games

In a multi-building system, one game is formulated at every time step. The first game of the day begins at $t = 1$, in which each building acts as an autonomous agent and determines the schedule of its power consumption for time window from $t=1$ to $t=N_p$. However, only the result of $t=1$ will be implemented. Similarly, the second game begins at $t=2$, followed by games repetitively until the last game is executed at the end of the day.

3.2 Players

In each game, a player is a building in the group. Specifically, the local supervisory controller of each building that determines the power consumption or temperature set point at every time interval is a player. The players are denoted as a finite set

$$N=\{1,2,...,n\}. \tag{5}$$

3.3 Actions

The action of each player determines a schedule of the amount of power consumption for the current sliding window timeframe. The action of player $i \in N$ in tth game is denoted as $a_i(t)$, then

$$a_i(t)=y_i(t)=[l_i(t)...l_i(t+N_p-1)]^{\mathrm{T}}, \tag{6}$$

where

$$l_i(t) \in L_i, \tag{7}$$

and L_{i} is set of all possible power consumption choices. The action set of player i is denoted as

$$A_i=\underbrace{L_i \times ... \times L_i}_{N_p}. \tag{8}$$

To make the action set A_i finite, the power consumption set is discretized using a minimum threshold $l_{i,\min}$, a maximum threshold $l_{i,\max}$, and a step size Δl_i so that

$$l_i=[l_{i,\min}:l_i:l_{i,\max}]. \tag{9}$$

The joint action set of all the players $N=\{1,2,...,n\}$ is denoted as

$$A=A_1 \times ... \times A_n. \tag{10}$$

An action profile of all the players is defined as

$$a(t)=\Big(a_1(t),a_2(t),...,a_n(t)\Big) \in A, \tag{11}$$

and a profile of the actions of all the players other than player i is defined as

$$a_{-i}(t)=\Big(a_1(t),...,a_{i-1}(t),\ a_{i+1}(t),...,a_n(t)\Big). \tag{12}$$

Therefore, $a(t)$ can also be rewritten as

$$a(t)=\Big(a_i(t),a_{-i}(t)\Big). \tag{13}$$

3.4 Payoff functions

The payoff function represents the benefit that a player can obtain as a result of an action profile, and all the players would like to maximize their payoffs.

In this paper, the payoff of each player is formulated and consists of three parts. Part 1 is the normalized operation cost penalty, which describes the economic benefit of a player. A low operation cost corresponds to a high payoff for each player. Part 2 is the temperature comfort level payoff, which is a reformulation of the temperature range constraint to describe the benefit gains from staying within or the loss from falling outside the comfortable temperature range. Part 3 is the total power load payoff reformulated from the total power load constraint to introduce the penalty for exceeding the desired maximum peak threshold.

These three aspects of the payoff function have different units and scales of dollar, kelvin, and watt. Thus, they are normalized into dimensionless quantities with the same scale of [0, 1].

Part 1: Normalized Operation Cost Penalty

$$u_{1i}\Big(l_i(t),p(t)\Big)=1-\frac{c_i\Big(l_i(t),p(t)\Big)}{c_{\max}},$$

where the second term in the equation is the current operation cost divided by the maximum possible operation cost, representing a normalized cost with scale [0,1]. However, since the higher the cost, the lower the payoff, thus to define payoff, we use 1 minus the term, the result is still a dimensionless quantity within [0, 1].

Part 2: Temperature Comfort Constraint Penalty

$$u_{2i}\Big(T_i(t-1),T_{\mathrm{OD}}(t),l_i(t)\Big)$$

$$=\frac{1}{1+\max\left\{0,\left(T_i(t)-\frac{T_{i\_lb}+T_{i\_ub}}{2}\right)^2-\left(\frac{T_{i\_ub}+T_{i\_lb}}{2}\right)^2\right\}},$$

where the term within max is the distance of the temperature from the middle point of T_{i_lb} and T_{i_ub} compared with the half distance $\frac{T_{i\_ub}-T_{i\_lb}}{2}$. If the temperature remains within the comfort zone specified by T_{i_lb} and T_{i_ub}, then the value would be 1; otherwise, it is a positive value less than 1. Moreover, this part is within [0, 1].

Part 3: Total Power Load Constraint Penalty

$$u_{3i}\Big(l_i(t)\Big)=\frac{1}{1+\max\left\{0,\sum_{i=1}^{n}l_i(t)-L_r\right\}},$$

where the term within max is the distance of the total power load beyond the constraint L_r. If the power load summation remains within the threshold, then the value would be 1; otherwise, it is a positive value less than 1. Moreover, this part is scaled within [0, 1].

Payoff formula: Weighted Average of Three Parts

$$u_i\Big(l_i(t),p(t),T_i(t-1),T_{\mathrm{OD}}(t)\Big) = \omega_1 u_{1i}\Big(l_i(t),p(t)\Big)+ \omega_2 u_{2i}\Big(T_i(t-1),T_{\mathrm{OD}}(t),l_i(t)\Big)+ \omega_3 u_{3i}\Big(l_i(t)\Big). \tag{14}$$

The overall payoff is formulated as a weighted average of the three aforementioned parts of the penalty. The weights ω_1, ω_2, ω_3 are in [0, 1] with their summation being equal to 1. Each weight represents the importance of the corresponding penalty term and can be adjusted based on different system design focus. Since each part of the penalty is a dimensionless quantity within [0, 1], the overall weighted average has the same scale.

3.5 Global payoff

The global utility/global payoff/social welfare of time slot t is usually defined as the summation of the payoffs of all the players at time slot t, that is,

$$W_t\Big(L(t),p(t),T(t-1),T_{\mathrm{OD}}(t)\Big) = \sum_{i\in\{1,\dots,n\}} u_i\Big(l_i(t),p(t),T_i(t-1),T_{\mathrm{OD}}(t)\Big).$$

The daily global utility is usually defined as

$$W = \sum_{t\in\{1,\dots,24\}} W_t.$$

Here to ensure that all payoffs and utilities have the same scale [0,1], we define the global utility of time slot t as the average payoff of all the players, that is,

$$W_t\Big(L(t),p(t),T(t-1),T_{\mathrm{OD}}(t)\Big) = \frac{1}{n}\sum_{i\in\{1,\dots,n\}} u_i\Big(l_i(t),p(t),T_i(t-1),T_{\mathrm{OD}}(t)\Big),$$

and the daily global utility as the average hourly global utility over a whole day

$$W(L,p,T,T_{\mathrm{OD}}) = \frac{1}{24}\cdot\frac{1}{n}\sum_{t\in\{1,\dots,24\}}\sum_{i\in\{1,\dots,n\}} u_i\Big(l_i(t),p(t),T_i(t-1),T_{\mathrm{OD}}(t)\Big).$$

3.6 Strategy: a payoff-based distributed learning algorithm

Step 1: Initialization. Each player is initialized with a state $[\overline{a}_i(t),\overline{u}_i(t),m_i(t)]$, where $\overline{a}_i(t)$ is the benchmark action, $\overline{u}_i(t)$ is the benchmark payoff, and $m_i(t)$ is the mood. Initially, $\overline{a}_i(t)$ is randomly selected from the action set, $\overline{u}_i(t)$ is zero, and $m_i(t)$ is discontent.

Step 2: Each player updates its action by using different probability functions depending on the mood of a player.

If $m_i(t)$ is content, then a player selects a new action according to the following probability distribution:

$$P_i^{a_i(t)} = \begin{cases} \dfrac{\varepsilon^c}{|A_i|-1} & \text{for} \quad a_i(t)\neq\overline{a}_i(t) \\ 1-\varepsilon^c & \text{for} \quad a_i(t)=\overline{a}_i(t), \end{cases} \tag{15}$$

where $c\geqslant n$ is a constant, and $0<\varepsilon<1$ is an exploration rate.

If $m_i(t)$ is discontent, then a player selects a new action according to the following probability distribution:

$$P_i^{a_i(t)} = \frac{1}{|A_i|} \quad \text{for every} \quad a_i(t)\in A_i. \tag{16}$$

Although uncommon in traditional game theory, mood $m_i(t)$ is a generic representation in distributed learning-based game theory. Mood is an internal state variable that determines the underlying status of a player. Usually, mood has two distinct modes, namely, content and discontent. When the mood is content, a player tends to stay at its current benchmark action with high probability. When the mood is discontent, a player tends to change its action by searching and selecting through the action set. Upon selecting a new action, the player updates its mood by evaluating the corresponding payoff. Usually, a high received payoff corresponds to a great probability that the mood is updated to be content.

Therefore, instead of directly selecting an action to maximize the payoff in traditional game theory, in distributed learning-based game theory, actions are implicitly adjusted by the payoff through the guidance of mood, which is the internal reflection of the payoff.

Step 3: Each player calculates its payoff according to payoff function (Eq. 14).

Step 4: Each player updates its mood.

If $m_i(t)$ is content and the action of the player remains the same, then the new mood is content;

If $m_i(t)$ is content but the action changes

$$[\overline{a}_i(t),\overline{u}_i(t),C] \xrightarrow{[a_i(t),u_i(t)]} \begin{cases} [a_i(t),u_i(t),C] \text{ with prob } \varepsilon^{1-u_i(t)} \\ [a_i(t),u_i(t),D] \text{ with prob } 1-\varepsilon^{1-u_i(t)} \end{cases}. \tag{17}$$

If $m_i(t)$ is discontent,

$$[\bar{a}_i(t),\bar{u}_i(t),D] \xrightarrow{[a_i(k),u_i(t)]}$$

$$\begin{cases} [a_i(t),u_i(t),C] \ \text{with} \ \ \text{prob} \ \ \varepsilon^{1-u_i(t)} \\ [a_i(t),u_i(t),D] \ \text{with} \ \ \text{prob} \ \ 1-\varepsilon^{1-u_i(t)} \end{cases}. \qquad (18)$$

Step 5: Global system convergence condition.

A state profile is defined as the joint states of all the players. The frequency of each state profile is counted. If the frequency of any state profile is larger than a predetermined threshold, such as 90%, then the game converges to a stable point in a global system and the algorithm is terminated. Otherwise, Step 2 is performed again, followed by the succeeding steps. The overall procedure of the proposed algorithm is summarized in the schematic shown in Fig. 1. The following theorem describes the property of the convergence point and states its global optimality.

Theorem 1. Let G be an interdependent n-person game on a finite joint action space. Under the distributed learning algorithm, given any probability $p < 1$, if the exploration rate ε is sufficiently small, then for all sufficiently large time t,

$$a \in \underset{a \in A}{\operatorname{argmax}}\, W(a) = \sum_{i \in N} u_i(a),$$

with at least probability p (Marden et al., 2014).

Theorem 1 serves as the guarantee of the convergence property under the state transition strategy specified by the aforementioned distributed learning algorithm. Under the assumption of interdependent n-person game and finite joint action space, the distributed algorithm can guarantee the probabilistic convergence of the action profile that can maximize the global payoff as long as the exploration rate is sufficiently small and the experiment time is sufficiently large.

Remark 1. With the problem of cooperative building temperature control and the discretization of power consumption space, such application satisfies the convergence criteria. However, such convergence is a probabilistic convergence as opposed to an almost

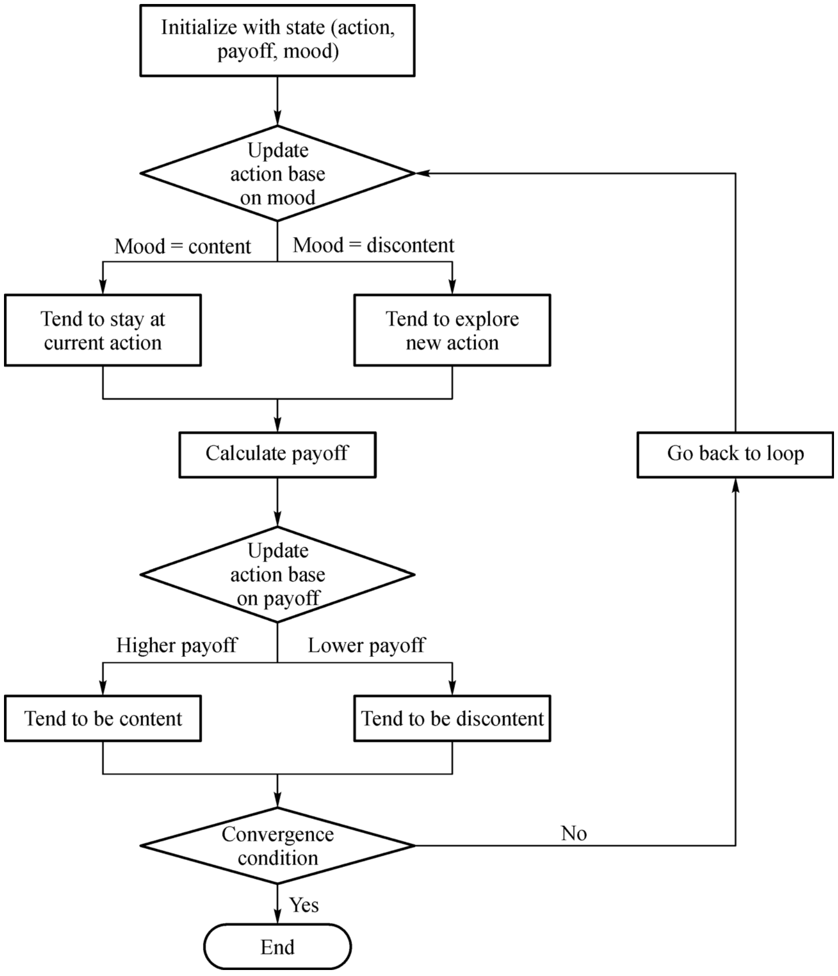

Fig. 1 Schematic of the payoff-based distributed learning algorithm

certain convergence. In practice, due to the limitation of exploration rate and sufficiently large experiment time, the distributed algorithm can achieve a global near-optimal solution, that is, the solution converges to the global optimum with a probability approaching one but never equal to one.

4 Case studies

In this section, numerical examples are provided to evaluate the performance of the proposed sliding window game formulation with a distributed learning strategy. For ease of illustration, a cooling scenario for a three-building energy optimization and temperature comfort cooperation problem is considered. For real-life consumers, these buildings can be interpreted as office buildings for a large corporation or classroom buildings for a university campus.

4.1 Hourly game

In this hourly game, one game is played at the beginning of every hour. In each game, the players focus only on the current time slot and try to maximize the global utility of that time slot following the strategy stated in the last section.

In each game, the action of a player is the amount of his energy consumption in the current time slot, that is,

$$a_i(t) \ = l_i(t).$$

Payoff can be calculated as

$$u_i\Big(l_i(t),p(t),T_i(t-1),T_{\mathrm{OD}}(t)\Big) = \omega_1\left[1-\frac{c_i\Big(l_i(t),p(t)\Big)}{c_{\max}}\right]+$$

$$\omega_2\left[\frac{1}{1+\max\left\{0,\left(T_i(t)-\frac{T_{i\_lb}+T_{i\_ub}}{2}\right)^2-\left(\frac{T_{i\_ub}-T_{i\_lb}}{2}\right)^2\right\}}\right]$$

$$+\omega_3\left[\frac{1}{1+\max\left\{0,\sum_{i=1}^{n}l_i(t)-L_r\right\}}\right],$$

and global utility is calculated as

$$W_t\Big(L(t),p(t),T(t-1),T_{\mathrm{OD}}(t)\Big)$$

$$=\frac{1}{n}\sum_{i\in\{1,\dots,n\}}u_i\Big(l_i(t),p(t),T_i(t-1),T_{\mathrm{OD}}(t)\Big).$$

By running the distributed learning algorithm described in the last section as the strategy of each player, the game can converge to a global optimal point, which is an efficient state profile that can achieve the maximum of the global utility of that certain time slot. After convergence, the optimal action profile is outputted to each corresponding high-level controller to implement control performance. The entire process is repeated every hour.

The following are several simulation results for hourly optimization. The red line in Figs. 2, 3, and 4 represents the optimal solution obtained through exploring the whole solution set by using a centralized manner, and the blue line represents the solution obtained by the distributed learning algorithm.

Figure 2 shows that the distributed learning algorithm can achieve a near-optimal global maximization solution. In this case, the distributed learning algorithm could obtain an average of 95.51% of the centralized optimal performance. The total utility in that figure consists of three components. The first part is the energy cost, which is a function of the hourly price and hourly energy consumption amount, and is shown in the three graphs in Fig. 3. The second part is the total energy consumption constraint for the three buildings described by a penalty function with threshold set as 1.5 kWh, which is illustrated in the third graph of Fig. 3. The third part is the temperature comfort level, which is described by a penalty function with the comfort temperature zone set as [67°F, 79°F], and is shown in Fig. 4.

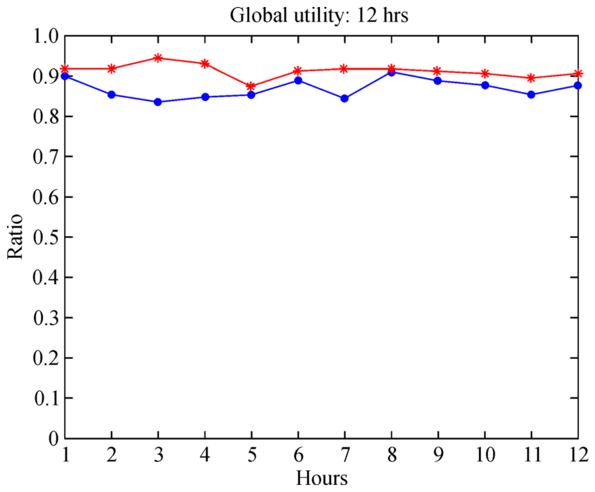

Fig. 2 Global utility distributed (blue dot) versus centralized optimum (red cross).

4.2 Sliding window game

In the sliding window game, one game is also played at the beginning of every hour. However, in each game, the players focus on an *h*-hour time horizon, attempting to optimize the global utility of that time period.

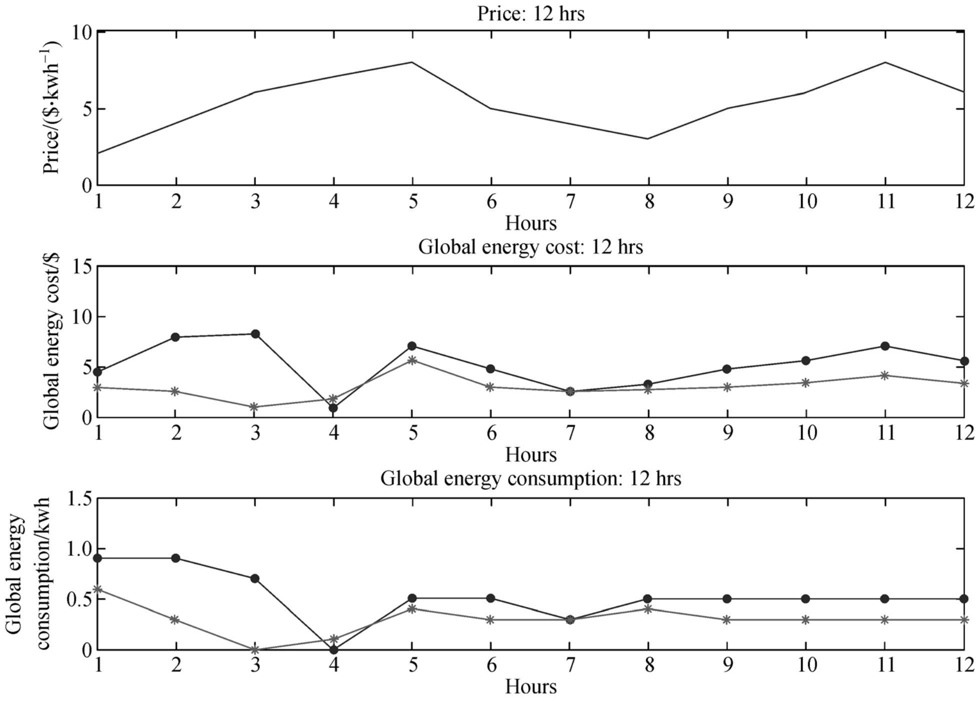

Fig. 3 Global utility components 1 and 2: Distributed (blue dot) versus centralized optimum (red cross)

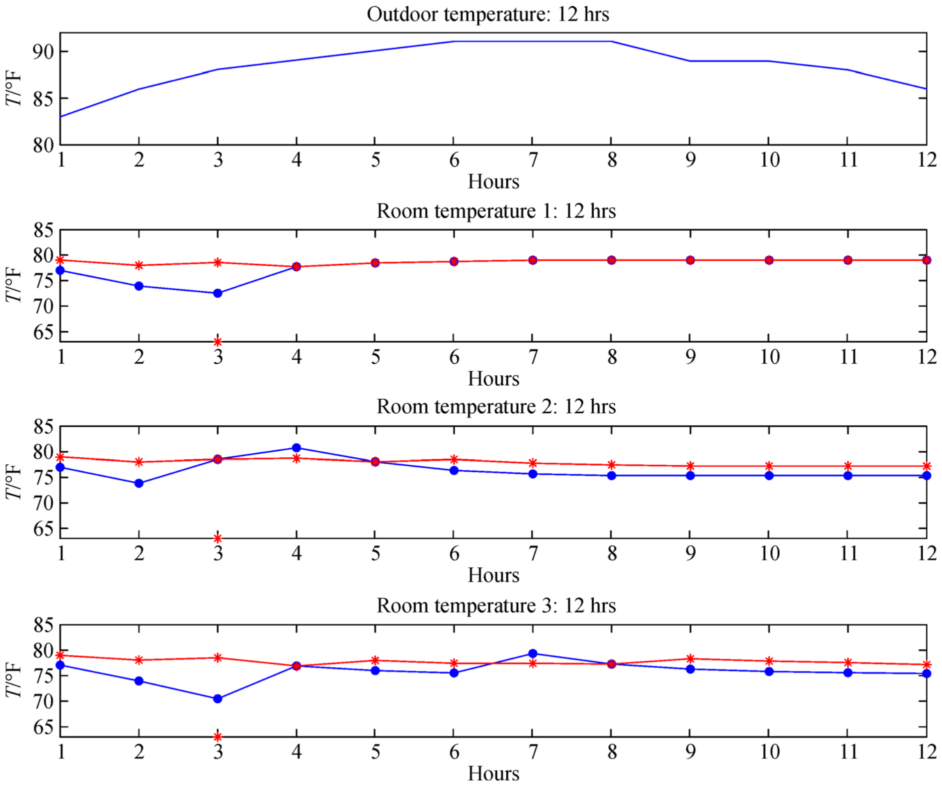

Fig. 4 Global utility component 3: Distributed (blue dot) versus centralized optimum (red cross)

In this case, the action of a player in each game is an energy consumption vector, that is,

$$a_i(t) = [l_i(t), l_i(t+1), \ldots, l_i(t+h-1)]^{\mathrm{T}}.$$

The payoff of a player can be calculated as

$$u_i\Big(l_i(t), p(t), T_i(t-1), T_{\mathrm{OD}}(t)\Big) = \omega_1\left[1 - \frac{c_i\Big(l_i(t), p(t)\Big)}{c_{\max}}\right]$$

$$+\omega_2\left[\frac{1}{1+\max\left\{0, \left(T_i(t) - \frac{T_{i\_lb}+T_{i\_ub}}{2}\right)^2 - \left(\frac{T_{i\_ub}-T_{i\_lb}}{2}\right)^2\right\}}\right]$$

$$+\omega_3\left[\frac{1}{1+\max\left\{0, \sum_{i=1}^{n} l_i(t) - L_r\right\}}\right],$$

$$u_i\Big(l_i(t\sim t+h-1),\ p(t\sim t+h-1),\ T_i(t-1\sim t+h-2),$$

$$T_{\mathrm{OD}}(t\sim t+h-1)\Big)$$

$$= \frac{1}{h}\sum_{k=t}^{t+h-1} u_i\Big(l_i(k), p(k), T_i(k-1), T_{\mathrm{OD}}(k)\Big),$$

and the global utility becomes

$$W_{t\sim t+h-1}\Big(L(t\sim t+h-1),\ p(t\sim t+h-1), T(t-1\sim t+h-2),$$

$$T_{\mathrm{OD}}(t\sim t+h-1)\Big)$$

$$= \frac{1}{nh}\sum_{i\in\{1,\ldots,n\}}\sum_{k=t}^{t+h-1} u_i\Big(l_i(k), p(k), T_i(k-1), T_{\mathrm{OD}}(k)\Big).$$

In this sliding window game, the global utility over a time period is maximized so that buildings can perform pre-cooling or pre-heating before the peak-price period.

Figure 5 shows a simulation comparison between the hourly game and the h-horizon game. The blue line represents the optimal solution attained by the hourly game, while the black line represents the game with a four-hour-long moving window. In a Monte Carlo simulation performed 100 times, the four-hour moving horizon game achieves a daily utility that is better by an average of 6.03% than the hourly game.

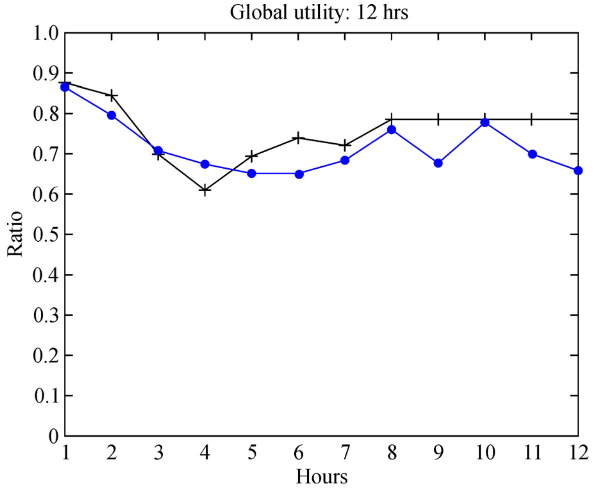

Fig. 5 Sliding window size = 4 (black cross) vs. window size = 1 (blue dot)

Figure 6 shows the room temperature of the three buildings under the h-horizon game and the corresponding distributed learning strategy. The black and red lines and the green circle represent the room and outdoor temperatures and the pre-cooling effect in off-peak hours, respectively. The pre-cooling effect is the result of the maximization of payoff for each building so that each building tends to consume more energy when the electricity price is lower. The pre-cooling effect is also the result of the coordination among buildings; such coordination causes the pre-cooling periods of the buildings to tend to offset one another.

5 Conclusions

In this paper, the distributed global utility maximization problem is formulated as a series of sliding window games in which each building is treated as a player. The global utility is decomposed to the payoff of each player, and each game is played over a control horizon from the current time step. The proposed distributed learning algorithm in game theory effectively ensures that each building can learn to play a part of the optimal solution. During the games, each player successfully maximizes its own payoff based on the action it played and the payoff it received without knowing others. Overall, the distributed algorithm achieves a near-optimal global solution, that is, the solution converges to the centralized optimal solution with a probability approaching one.

The effectiveness of this method is demonstrated through a simulation of three building systems. The convergence property of the methodology is successfully demonstrated. Moreover, pre-heating/cooling effect during off-peak period and autonomous heating/cooling discharge during on-peak period are observed because the method uses sliding window games with a prediction horizon.

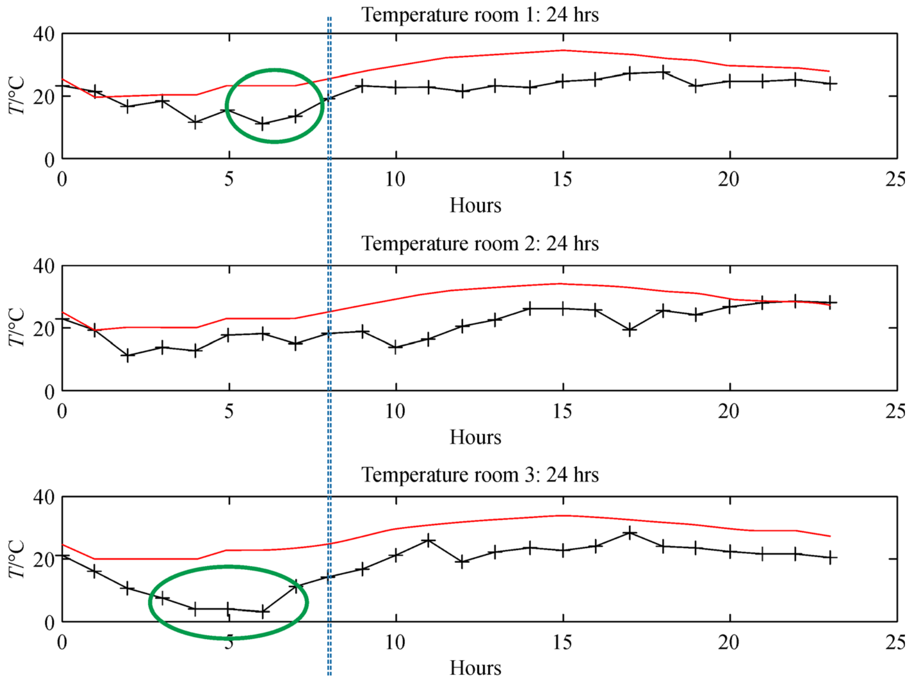

Fig. 6 Outdoor (red), indoor (black cross) temperature, and pre-cooling

Acknowledgements This work is supported in part by the Fundamental Research Program of the Shenzhen Committee on Science and Innovations (Ji20160207), the Texas–Wisconsin–California Control Consortium, and Alberta Innovates Technology Futures Postdoctoral Fellowship Program at the University of Alberta.

References

Chai B, Chen J, Yang Z, Zhang Y (2014). Demand response management with multiple utility companies: a two-level game approach. IEEE Transactions on Smart Grid, 5(2): 722–731

Deng R, Yang Z, Chen J, Asr N R, Chow M Y (2014). Residential energy consumption scheduling: a coupled-constraint game approach. IEEE Transactions on Smart Grid, 5(3): 1340–1350

Deng R, Yang Z, Chow M Y, Chen J (2015). A survey on demand response in smart grids: mathematical models and approaches. IEEE Transactions on Industrial Informatics, 11(3): 570–582

Deng R, Zhang Z, Ren J, Liang H (2016). Indoor temperature control of cost-effective smart buildings via real-time smart grid communications. In: Global Communications Conference (GLOBECOM), 2016 IEEE. IEEE, 1–6

Dimarogonas D V, Frazzoli E, Johansson K H (2012). Distributed event-triggered control for multi-agent systems. IEEE Transactions on Automatic Control, 57(5): 1291–1297

Dounis A, Manolakis D (2001). Design of a fuzzy system for living space thermal-comfort regulation. Applied Energy, 69(2): 119–144

Dounis A I, Caraiscos C (2009). Advanced control systems engineering for energy and comfort management in a building environment—a review. Renewable & Sustainable Energy Reviews, 13(6–7): 1246–1261

Fan Y, Feng G, Wang Y, Song C (2013). Distributed event-triggered control of multi-agent systems with combinational measurements. Automatica, 49(2): 671–675

Forouzandehmehr N, Perlaza S M, Zhu H, Poor H V (2013). A satisfaction game for heating, ventilation and air conditioning control of smart buildings. In: Proc. IEEE GLOBECOM, 3164–3169

Kolokotsa D, Stavrakakis G, Kalaitzakis K, Agoris D (2002). Genetic algorithms optimized fuzzy controller for the indoor environmental management in buildings implemented using PLC and local operating networks. Engineering Applications of Artificial Intelligence, 15(5): 417–428

Kummert M, André P, Nicolas J (2001). Optimal heating control in a passive solar commercial building. Solar Energy, 69: 103–116

Levermore G J (2000). Building Energy Management Systems: Applications to Low-Energy HVAC and Natural Ventilation Control. Oxfordshire: Taylor & Francis

Li N, Chen L, Low S H (2011). Optimal demand response based on utility maximization in power networks. In: 2011 IEEE Power and Energy Society General Meeting. IEEE, Piscataway, 1–8

Li Z, Ren W, Liu X, Fu M (2013). Distributed containment control of multi-agent systems with general linear dynamics in the presence of multiple leaders. International Journal of Robust and Nonlinear Control, 23(5): 534–547

Ma J, Qin J, Salsbury T, Xu P (2012). Demand reduction in building energy systems based on economic model predictive control. Chemical Engineering Science, 67(1): 92–100

Ma J, Qin S J, Li B, Salsbury T (2011). Economic model predictive control for building energy systems. In: Innovative Smart Grid Technologies, 2011 IEEE PES. IEEE, 1–6

Ma J, Qin S J, Salsbury T (2014). Application of economic MPC to the energy and demand minimization of a commercial building. Journal of Process Control, 24(8): 1282–1291

Marden J R, Young H P, Pao L Y (2014). Achieving Pareto optimality through distributed learning. SIAM Journal on Control and Optimization, 52(5): 2753–2770

McCartney K J, Nicol J F (2002). Developing an adaptive control algorithm for Europe. Energy and Buildings, 34: 623–635

Mohsenian-Rad A H, Wong V W, Jatskevich J, Schober R, Leon-Garcia A (2010). Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid. IEEE Transactions on Smart Grid, 1(3): 320–331

Moon J W, Kim J J (2010). Ann-based thermal control models for residential buildings. Building and Environment, 45(7): 1612–1625

Oldewurtel F, Parisio A, Jones C N, Morari M, Gyalistras D, Gwerder M, Stauch V, Lehmann B, Wirth K (2010). Energy efficient building climate control using stochastic model predictive control and weather predictions. In: American Control Conference (ACC), 58(8): 5100–5105

Shaikh P H, Nor N B M, Nallagownden P, Elamvazuthi I, Ibrahim T (2014). A review on optimized control systems for building energy and comfort management of smart sustainable buildings. Renewable & Sustainable Energy Reviews, 34: 409–429

Široký J, Oldewurtel F, Cigler J, Prívara S (2011). Experimental analysis of model predictive control for an energy efficient building heating system. Applied Energy, 88(9): 3079–3087

Wang S, Ma Z (2008). Supervisory and optimal control of building HVAC systems: a review. HVAC & R Research, 14(1): 3–32

Weng T, Agarwal Y (2012). From buildings to smart building-sensing and actuation to improve energy efficiency. IEEE Design & Test of Computers, 29(4): 36–44

Wright J A, Loosemore H A, Farmani R (2002). Optimization of building thermal design and control by multi-criterion genetic algorithm. Energy and Buildings, 34: 959–972

Zaheer-Uddin M, Zheng G (2000). Optimal control of time-scheduled heating, ventilating and air conditioning processes in buildings. Energy Conversion and Management, 41(1): 49–60

Zhang Z, Deng R, Yuan T, Qin S J (2017). Distributed optimization of multi-building energy systems with spatially and temporally coupled constraints. In: American Control Conference (ACC), IEEE, 2913–2918

Zhang Z, Deng R, Yuan T, Joe Qin S (2016). Bi-level demand response game with information sharing among consumers. IFAC-PapersOn-Line, 49(7): 663–668

Zhang Z, Li G, Salsbury T, Qin S J (2014). 389842 Game theory based distributed temperature control for energy saving of smart buildings. In: 2014 AICHE Annual Meeting, Atlanta, GA