Received June 2, 2020, accepted June 17, 2020, date of publication June 22, 2020, date of current version July 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3004151

Real-Time Scheduling of Operational Time for Smart Home Appliances Based on Reinforcement Learning

MURAD KHAN, JUNHO SEO, AND DONGKYUN KIM, (Member, IEEE)
School of Computer Science and Engineering, Kyungpook National University, Daegu 41566, South Korea

Corresponding author: Dongkyun Kim (dongkyun@knu.ac.kr)

This work was supported in part by the BK21 Plus Project (SW Human Resource Development Program for Supporting Smart Life) funded by the Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, South Korea, under Grant 21A20131600005, and in part by the Korea Research Fellowship Program funded by the Ministry of Science and ICT through the National Research Foundation of Korea under Grant 2019H1D3A1A01102987.

ABSTRACT The scheduling of the operational time of household appliances requires several parameters to be tuned according to the available energy supplied to a smart home. However, scheduling of operational time of multiple appliances in a smart home itself is the NP-hard problem and thus requires an intelligent, heuristic method to be solved in polynomial time. In this research work, we propose Real-time Scheduling of Operational Time of Household Appliances based on the well-known value iterative reinforcement learning called Quality learning (RSOTHA-QL). The proposed RSOTHA-QL scheme operates in two phases. In the first phase, the agents of the Q learning act by interacting with the smart home environment and obtain a reward. The reward value is further utilized to schedule the operational time of household appliances in the next state ensuring minimum energy consumption. In the second phase, the dissatisfaction arises due to scheduling of operational time of the household appliances of the home user is maintained by categorizing the household appliances into three groups: 1) deferrable, 2) non-deferrable, and 3) controllable. Besides, using the shared memory synchronization phenomenon, the agents attached to each appliance of the smart home become coordinated. The simulation and experiments are performed in a smart home scenario comprised of a single user and multiple appliances. As compared with our previous research work using the Least Slack Time (LST) scheduling algorithm and scheduling based on demand-response strategy, it is revealed that the operational time of the household appliances is efficiently scheduled to reduce the energy consumption and dissatisfaction level of the home users significantly.

INDEX TERMS Energy consumption, Q learning, scheduling, smart home.

LIST OF ABBREVIATIONS

$E^{A}_{n,t}$	The energy of an appliance n belongs to category A at a time t
$E^{B}_{n,t}$	The energy of an appliance n belongs to category B at a time t
$E^{C}_{n,t}$	The energy of an appliance n belongs to category C at a time t
t	An hour time instance in 24 hours of a day
T	Total hours in a day i.e. 24
N	Total number of appliances
H	Total number of hours in a day
$\mathbf{T_{ON}}$	Turning on an electronic appliance
$\mathbf{T_{OFF}}$	Turning of an electronic appliance
$\mathbf{OV_{n,t}}$	A Boolean variable representing the operation of an electronic appliance *n* is either ON or OFF
PL	Operating power levels of an appliance
s	Represents the state of an appliance n
a	Represents the action performed by a Q-learning agent
r	Represents the reward matrix
$\mathbf{P_{ni}}$	The probability of interaction of a human *i* with an appliance n in a week
G	A goal state i.e. performing all actions with less amount of energy

The associate editor coordinating the review of this manuscript and approving it for publication was Sanjeevikumar Padmanaban.

max(E)	Maximum energy currently consuming at a smart home
v	An optimal policy for each iteration Q-learning algorithm
max(CL)	Represents maximum comfort level
min(EC)	Represents minimum energy consumption
$\boldsymbol{\beta_n}$	The priority of an appliance n
$\mathbf{m_n}$	The median level of energy demand
$\boldsymbol{\theta}$	Learning rate of Q-learning algorithm
$\boldsymbol{\gamma}$	A discounted factor of Q-learning algorithm

I. INTRODUCTION

The introduction of smart grid technologies addresses most of the energy wastage problems by efficiently utilizing the advantages of communication technologies. However, over time, smart grids based on traditional concepts and communication technologies exhibit many challenges such as inappropriate and delayed two-way communication between the grid and smart meter, pricing control based on demand-response (DR) systems, difficulties in communication with renewable energy sources, unevenness load scheduling, and management, etc. These challenges are addressed with the introduction of a new technology called the micro-grids. The micro-grids enhance the communication between the smart meters, load scheduling, etc. incorporating the distributed power generation systems, distributed energy storage systems, and controlling a small scale of consumers in real-time. Similarly, micro-grids can be installed in those locations where there is no access to the main power grids. Thus, micro-grids provide an easy solution with a minimum cost of installation and managing the power generation and load scheduling of the small-scale community. However, it is still challenging to control Home Energy Management Systems (HEMS) and the scheduling of household appliances using micro-grid technologies. Similarly, controlling the HEMS with micro-grids can increase the burden of processing huge amounts of data and optimal scheduling of operational time of household appliances in real-time. Thus, a solution is needed to overcome the excessive energy wastage due to the inappropriate scheduling of operational time of household appliances at the consumer end. Recently, the researchers developed several HEMS based on micro-grids [1]–[3] to efficiently address the inappropriate scheduling of operational time of household appliances. Most of the HEMS works on the principle of controlling appliances in the context of operational time with peak, semi-peak, and off-peak times. Thus, scheduling the devices in those time slots required a real-time connection with the micro-grids.

Initially, the HEMS systems were developed to schedule the household appliances autonomously using the DR systems. For instance, the HEMS systems used to schedule and control the household appliances' load profile based on the Time of Use (ToU) phenomenon [4], [5]. Considering ToU as a parameter for scheduling the appliances and managing the load in the smart home does not guarantee a well-developed system for homes with multiple appliances and multi-users. The applications of ToU based systems can help in scenarios where the requirements of the system are known before implementing the system. For instance, a smart home user can define an optimal schedule of operating appliances if a user knows the price information of the billing a day or an hour ahead. Similarly, the HEMS systems can reschedule appliances from peak or semi-peak time to the off-peak time. However, such systems are always depending on the intervention of a home user, weather conditions, and operating time of appliances. Also, if there are multiple users in a single home, then managing the working of HEMS systems may result in inappropriate scheduling of appliances. For instance, consider a scenario where one the smart home user entering the home from outside hot temperature and a user is already available at home. The user coming from outside hot temperature may switch on the air-conditioner with a high-power level while at the same time the other user available inside the home may want to operate the AC on low power level. Similarly, such scheduling techniques do not consider the behavior or interaction of a smart home user with household appliances.

To optimize the functionality of the existing HEMS systems, two-way communication is required between the HMES and the smart grid. Similarly, to support such two-way communication between the HEMS and smart grid, recently, researchers use various machine learning techniques [6]–[9]. The machine learning techniques give accurate results in situations where the data between the smart homes and smart grids continuously exchanging. Also, powerful machines are required to process the data on smart and micro-grids. However, if we think realistically, most of the time, it is difficult for smart grids to handle such a huge amount of data from a hundred and thousands of homes. Also, these homes are further consisting of several appliances and users. Thus, shifting the data analytics from smart homes to smart grids may result in inadequate outcomes. However, if the functionality of the smart grid is somehow divided into multiple edge nodes available at edges of communities, then the problem can be somehow resolved. Such a phenomenon is rarely discussed in the recent literature. However, these systems still need the support of smart metering and smart home edge networking. The smart meters support the modernization of the smart grids and two-way communication between the smart homes and the smart grids. However, the smart meter only communicates the energy values with the smart grids. On the other hand, the smart grid responds to the smart meters with the required amount of energy forming a smart grid pyramid concept as shown in Figure 1 [10].

To schedule the operation of the household appliances in smart homes, we propose a scheduling scheme for the operational time of smart home appliances based on reinforcement learning. Specifically, we want a system that does not force a home user to schedule the appliances based on a day or hour-ahead prices from the utilities. The proposed

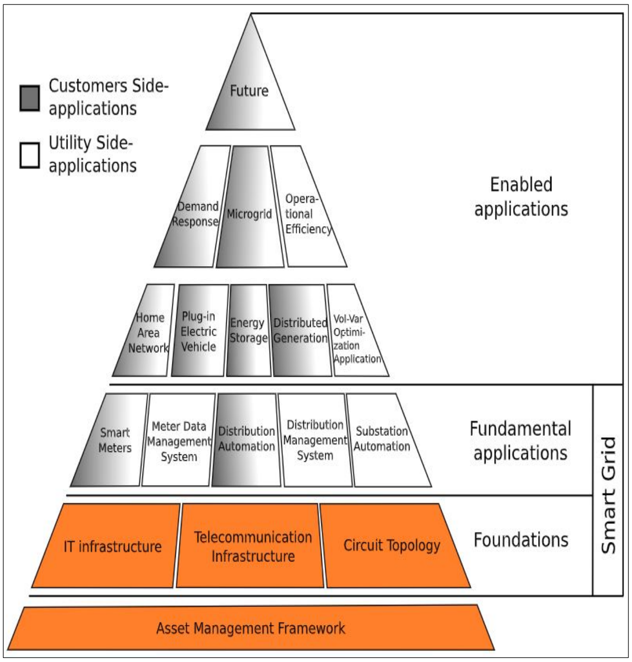

FIGURE 1. A smart grid pyramid [10].

scheme schedules the appliances within the smart home considering the currently available energy resources. A fixed amount of energy is considered for each time state in the smart home. Whenever the energy consumption of the smart home exceeds the fixed energy threshold, the proposed system switches off an appliance with less priority or change its power level from high to low.

To prioritize, the home appliances are grouped as deferrable, non-deferrable, and controllable as explained in Section II. The human behavior of turning on and off appliances at home at various times of the day is first modeled with the help of Q-Learning. The agents of the Q-learning are trained in a way to keep the limits of the energy intact and decide to operate the appliances based on the energy consumption of the appliances at a particular time of the day. For instance, a washing machine is always scheduled in off-peak times of the day (from 8:00 PM to 12:00 AM). Similarly, among two high energy consumable appliances, one of the appliances is always running at a low power level. Finally, the proposed system is easy to implement in the HEMS systems and supports a two-way communication architecture with the smart grid or an edge node available at the edge of a community.

The remaining of the paper is ordered as follows, Section II presents the related literature. Section III presents a discussion on the appliances used in a generic smart home and the scheduling techniques used to schedule these household appliances. Section IV outlines the problem formulation and the architecture of the proposed architecture. The discussion on results and experimental analysis are discussed in Section V. Finally, the conclusion is given in Section VI.

II. RELATED WORK

The future smart home will be consisting of many sensors controlling various activities ranging from security to maintaining a fair usage of energy [11]–[13]. The human behavior and interaction with smart home appliances changes rapidly and, therefore, relying on a fixed model for scheduling household appliances is difficult. In this regard, several research studies are carried out to design a generic recommendation system for smart home users based on machine learning techniques. Further, the recommendation system does provide the necessary combination, however, such systems perform better in item selection such as clothes, food, etc [14]. Therefore, apart from recommendation systems for smart home appliances scheduling, in the last decade, the researchers tried several different models ranging from traditional mathematical approaches to machine learning for scheduling the household appliances. However, many of them optimize the energy consumption of smart homes with the addition of photovoltaic (PV) systems. Most of the energy demands during peak hour time are fulfilled with the integration of renewable energy sources. Also, the main target of such approaches to reduces the electricity cost, however, less focus has given to energy consumption. A scheme discusses a similar approach to minimizing electricity costs is presented in [15]. To fulfill the energy demand, a PV panel is used as a microgrid. Further, an optimization algorithm is proposed to schedule smart home appliances based on Mixed-Integer Linear Programming (MILP). The proposed scheme scheduled the smart home appliances; however, the integration of the photovoltaic panel and the use of mixed-integer linear programming increases the complexity and cost of the deployment. Similar approaches for smart home appliances scheduling is proposed in [16], [17]. The proposed approach also adopted the flexibility of MILP to schedule the smart home appliances and the energy demand is fulfilled with the help of installing a PV system. Again, the discomfort level is reduced, however, the cost of deployment increases with the addition of PV systems. Also, the appliances are scheduled with fixed time slots and a high priority or uninterruptible devices are scheduled for continued operation.

The DR based system is used to handle the energy demands for smart homes from smart grids or micro-grids. Traditionally, the DR based approaches are used to encourage the home user to participate in maintaining the loads manually. However, over time the DR systems are automated with the help of smart grids optimization. An approach uses the DR system for energy optimization in smart grids is discussed in [18]. An energy hub is designed to control the flow of energies from various energy carriers resulting in an integrated DR system. An equilibrium state among energy hubs is achieved using the Nash equilibrium system. Also, the load shifting is achieved while interacting with the energy hub and shifting the load during peak hours. However, shifting a load during load time can increase the discomfort of the home user. Hour-ahead energy prices-based models are used to adjust the load among the household appliances with minimum electricity cost. However, such systems need continuous checking of the energy prices and maintaining the load among the household appliances. Resulting, in maximizing the smart home user discomfort level. An approach is used to somehow minimize the smart home user discomfort while using the hour ahead pricing model. The proposed system is integrated with a variety of household appliances, energy storage units, and distributed energy management systems. The historic data and weather information is used to predict the 24 hours ahead of energy prices. The main objective of this system is to present a home user with several choices for choosing the one better satisfying their energy demands and cost of the electricity. Also, each household appliance is assigned with an important parameter to prioritize the appliances. Thus, whenever an appliance of high importance is turned on, the energy demands are fulfilled from the PV system or storage unit. However, again the renewable energy sources are used to fulfill the energy demands. Also, switching the appliance to another state may increase the discomfort of the home user.

Machine learning techniques are recently adopted to consider the smart home user choice of energy demands with HEMS. However, employing machine learning techniques in HEMS systems require extensive training of the models with historic smart homes energy consumption data. Also, the machine learning systems need proper management and efficient interfacing with the smart home appliances via a Home Edge Computing (HEC) system, etc. Moreover, the microgrid needs to be tuned to provide seamless energy supply to the HEC systems and agents. However, the power supply required enough information on the HEC systems, the nature of the tasks, and the supply resources i.e. renewable, non-renewable, and storage. Thus, it becomes a challenging job for the micro-grid to deal with the uncertainty and irregularity of renewable sources due to many reasons such as weather conditions, accidental damage, etc. The nature of the problem of supplying energy from renewable sources makes it an NP-hard problem and thus needs an optimized solution. Though, it is difficult to model energy supply from renewable sources to the HEC networks. However, a system is proposed in [19] with an idea of solving the aforementioned problem by decomposing it in energy-efficient tasks assignment and energy supply plan into a Markov Decision Process (MDP). The first part is solved for each base station with a distributed approach, while the second part is solved with the deep reinforcement learning. The system is then tested with sending unpredictable tasks to the proposed system and the proposed system converts it to energy-efficient task assignment by assigning a particular amount of energy to each task. Secondly, the tasks are handled efficiently using the deep reinforcement learning algorithm which incorporates the present states of the system and results in reduced power consumption for each task. However, the proposed system does not provide the results of how much time the proposed deep reinforcement algorithms require for training and computation of the energy supplied to the incoming tasks. Further, the prediction model of predicting future energy requires insight analysis of the system prediction accuracy. The demand for energy is increasing every day and thus increasing its cost for a generation. A number of approaches based on Time-of-use (ToU) have been proposed to schedule the home appliances to achieve low cost and optimize the demand for electricity. However, the ToU schemes have a number of problems such as scheduling of appliances that are always based on traditional algorithms which are sometimes produced inappropriate results in generic scenarios. Similarly, the demand response methods which are implemented recently are mainly based on mixed-integer linear programming. However, these methods are tested in a day ahead prices which again results in increasing the cost compared to hour-ahead systems. Similarly, optimizing the DR systems is a challenging job due to the tuning of several parameters and variables. A solution is presented based on virtualizing DR systems considering the energy management system as a virtual retailer (leader) offering virtual retail prices the home users (follower). The followers are supposed to follow the instructions from the leader which results in an optimization problem solved by incorporating the Stackelberg game theorem. The Stackelberg game theorem finds the optimal policy for the follower. The leader is programmed to provide the required energy needed by the follower. The results show that the Stackelberg game theorem optimizes the load among the available followers while controlling the real-time prices of the leader. Although the Stackelberg game theorem is an efficient solution to handle 1-to-N-follower DR energy problems. However, the interaction of the followers with the appliances is neglected in the proposed scheme which results in increasing the energy demand among the available smart homes attached to a leader. Thus, a machine-learning algorithm could be used to solve such limitations of the proposed scheme.

III. CHARACTERISTICS AND BACKGROUND DISCUSSION ON ELECTRONIC APPLIANCES

In general, a smart home consists of several appliances ranging from low to high energy consumption. Also, some of the appliances are always in the on state such as refrigerator, light bulbs at some particular time, television, etc. and some of the appliances required a specific task to complete at a specific time of the day or week. To differentiate between these appliances, we categorized these appliances into three categories as deferrable, non-deferrable, and controllable as follows.

A. CATEGORY A: NON-DEFERRABLE HOME APPLIANCES

These appliances cannot be shifted or scheduled into another time of the day or week. For example, a refrigerator belongs to this category. Since it is important for a refrigerator to remains in switched on mode all the time to keep the food items and other things from spoiling. Similarly, a television

(TV) is also belonging to the same category. The operation of the TV cannot be postponed to another time of the day or week. Also, all those appliances which belong to this category can increase the discomfort of the user if they are scheduled to another time of the day. These appliances should always be provided with the required amount of energy denoted with $E^{A}_{n,t}$ to keep the operation in running all the time where $t \in [1, 2, 3, \cdots, 24]$ hours of the day. Also, the user can perform only an "ON" action on such appliances. The energy consumption of a refrigerator is shown in Figure 2 [20]. The Figure shows that the energy consumption is high as the refrigerator is switched on. However, as soon as the required temperature reached, the refrigerator started working with lower energy consumption.

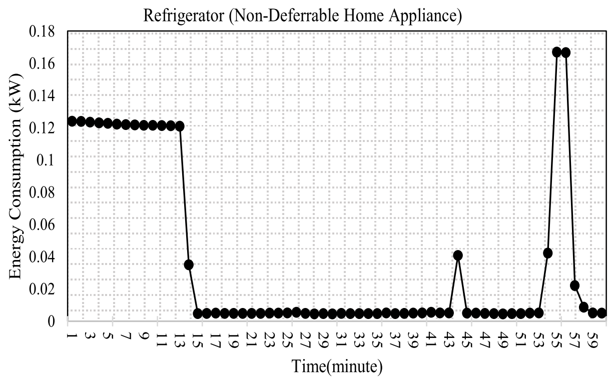

FIGURE 2. The energy consumption of a refrigerator.

B. CATEGORY B: DEFERRABLE HOME APPLIANCES

Such appliances can be switched or scheduled to another time of the day where the requirements of the energy are less. For instance, if the energy of the home is exceeding the limit of the energy, an appliance from this category can be switched to another time of the day or can be switched off. However, once an appliance of such a category is switched on, it cannot be halted before the operation completed. If such appliances are halted during the operational time, this will increase the discomfort of the user. Also, the users can perform only two types of actions on such appliances i.e. "ON" and "OFF". The energy consumption depends on the time slots in which they are operating. Also, shifting such appliances to another time slot of the day increases the discomfort level of the user. For instance, if an appliance of such category is shifted from time t to $t + 3$ where t represents an hour then this means that the user should wait for three hours to operate the appliance. So, along with energy consumption, these appliances also increase the dissatisfaction level of the user. The energy consumption of such appliances represents with $E^{B}_{n,t} = OV_{n,t} \times e^{B}_{n,t}$ where OV is a Boolean variable representing the operation is either on or off. For example, a washing machine belongs to such a category. The smart home user always wants to schedule a washing machine in off-peak time. Figure 3, shows the energy consumption of a dishwasher [20]. A dishwasher belongs to a category of deferrable devices and it requires approximately an hour and 45 minutes to complete the task of washing dishes and dry it.

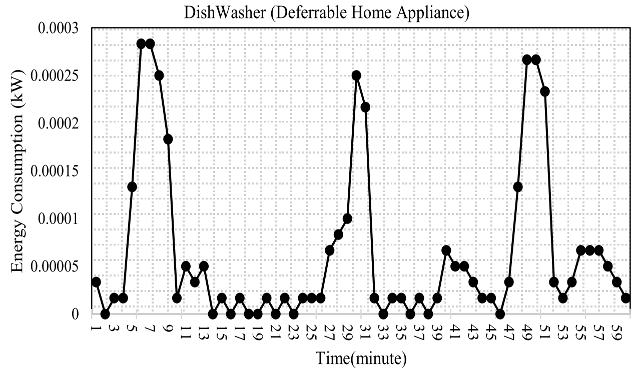

FIGURE 3. Energy consumption of a dishwasher.

C. CATEGORY C: CONTROLLABLE HOME APPLIANCES

Apart from the above two categories, other appliances can be operated over different power ratings. For instance, an AC can be operated on different power levels depending on the temperature requirements of an environment. Also, changing the power levels of such devices may results in increasing the discomfort level of the user, however, it can reduce energy consumption. Also, a number of power levels can be defined for such appliances. In this study, we defined five different power levels i.e. $PL = [PL_1, PL_2, PL_3, PL_4, PL_5]$ for these appliances where PL_1 represents a high power level and PL_5 represents low power level. In this category of appliances, a user can perform seven different actions five for changing power levels and two for switching on or off an appliance. The total energy consumption of such appliances i.e. N depends on the power requirements of the appliances at a time t i.e. $E^{C}_{n,t} = \sum_{n=0}^{N} \sum_{PL_1}^{PL_5} e^{C}_{n,t}$. The energy consumption of the AC for an amount of 1 minute time is shown in Figure 4 [21]. The graph reveals that energy consumption depends on the weather conditions of the smart home. For instance, if the weather is hot inside the home, the AC will require high energy to balance the temperature.

D. SMART HOME APPLIANCES SCHEDULING: A SINGLE USER SCENARIO

A single smart home scenario is extensively studied during the last decade. In a single user smart home scenario with several appliances, the scheduling is mainly based on the time of operation of an appliance. In general, the entire day is divided into time slots based on hours in a day. Further, the time of operation reserved for an appliance is divided based on on-peak and off-peak hours. For instance, two appliances requiring high energy cannot be switched on together. Similarly, some of the appliances such as AC require high energy, however, an AC can be switched to a low power level to run multiple appliances at the same time. Let assume the time horizon is divided into H hours,

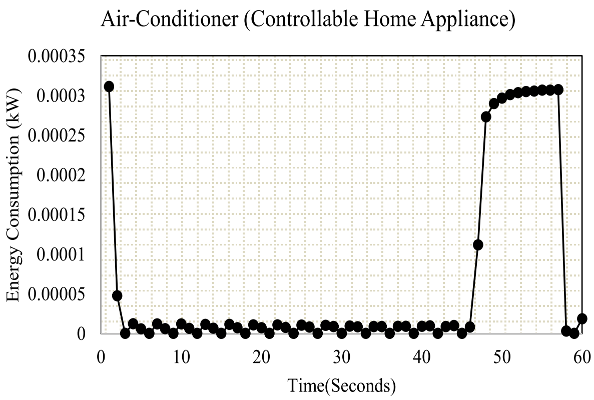

FIGURE 4. **Energy consumption of an air-conditioner.**

where H represent a positive integer between 1 and 24 i.e. $H = [h_1, h_2, h_3, \cdots h_{24}]$. Such an assumption or a similar approach is adopted in several scheduling algorithms based on dynamic programming, least slack time, game-theoretic model, and even machine learning [5], [22]–[27]. Also, such systems depend on several constraints such as starting time, ending time, time of operation, the total amount of energy to consume, etc. Thus, based on such constraints a model can be developed to efficiently distribute the available power among the household appliances. However, integrating many constraints can also lead to inappropriate scheduling in the case of a multi-user smart home scenario. In one of our previous studies, we used the LST scheduling algorithm to schedule the household appliances. One of the major problems in using LST is that it always does the current job and never check the future tasks. The scheduling is modeled with prior information of all the required variables for the LST algorithm. The LST then performs those tasks one by one. However, such weakness of the LST is handled in the same research by incorporating the sleep or inactive state after the occurrence of an interrupt or emergent task. Also, setting the priority of the appliances may somehow resolve the problem of handling emergent tasks. However, such a concept cannot be added to the LST as it works on the current jobs and does not investigate the future. The working of the LST for n appliances and a single user scenario is shown in Algorithm 1.

The LST algorithm always schedules an appliance if the appliance is operational and the energy required by the appliance does not exceed the limit of the maximum energy. For instance, if an appliance belongs to high priority appliances i.e. refrigerator, and if the appliance is turned on, its energy would exceed the limit. In such cases, the LST will first run the existing tasks and will not turn on the appliance until another appliance is switched off. In such situations, the LST always performs inappropriately, because the refrigerator must be turned on to avoid food spoiling. Similarly, the literature consisted of several scheduling algorithms based on the DR concept. These algorithms work on the principle of fulfilling the demand from a power grid

Algorithm 1 LST Based Home Appliances Scheduling

1 **Input:** Requests and power consumption of appliances(p)
2 **Output**: optimal scheduling of appliances
3 Initialize R ← collection of requests
4 $k \in \mathbf{R}$
5 p(k) ← power consumption of an individual appliance ***k***
6 p_total ← Total power consumption of accepted requests
7 Max_Thr_Power ← Maximum power consumption
8 p = **0**
9 $\forall k \in R$ **do**
10 Check if k is non-preemptive and operational
11 accepted_requests ← ***k***
12 R −= ***k***
13 p_total + = **p(k)**
14 check if p_total ≤
15 Max_Thr_Power && k is operational
16 goto step **13**
17 end if
18 end for

whenever a new appliance is turned on. However, in such cases, the energy consumption may exceed the limit, or they do not consider the preference of appliances. Recently, several schemes presented with concepts of incorporating the appliances' priorities [28], [29]. However, assigning appliances priorities and load scheduling is a complex job due to the stochastic behavior of humans. Thus, relying on a system with a fixed infrastructure required a powerful real-time decision-making system. However, up to this day, none of the systems is designed to accurately model human behavior. Further, succeeding in some specific scenarios always help in developing more accurate and reliable systems [30]. In the case of operating home appliances and based on the past interaction of the humans with the household appliances, a system can be designed to accurately operate the home appliances via an intelligent agent. The agent can be programmed to somehow modeled the human behavior patterns and come up with actions like those performed by a home user. Considering a smart home as Markov-Decision Process (MDP) based environment where the human goes from one state to another state. For example, a home user always switching on lights of the home from 6:00 PM onwards every day. If such behavior of the user is modeled into an agent, the agent can perform the same action of switching off the lights at time 6:00 PM. However, while switching on the lights, the home user never considered the total energy consumption at 6:00 PM. Let assume, switching on the lights increases the total energy from the maximum power consumption of the home. In such cases, the agents can program to check the entire home energy and rescheduled other appliances or changing the power level from high to low to consume energy within the energy consumption threshold. Machine learning-based scheduling of home appliances is extensively studied in the recent literature due to their advantages of making real-time decisions [31]–[33]. However, these systems still need further improvements to optimally scheduled the household appliances considering the human appliance interaction. Therefore, in this research work, we come up with an intelligent scheduling system based on Q learning to enhance the working of our previous study presented in [5].

IV. REAL-TIME SCHEDULING OF OPERATIONAL TIME OF HOUSEHOLD APPLIANCES USING QL

In this section, the proposed RSOTHA-QL scheme is modeled and presented considering a home scenario with a single user and multiple appliances. Also, the explanation of actions, states, and rewards is shown based on reinforcement learning.

A. PROBLEM FORMULATION

1) MOTIVATION

The energy consumption of a smart home entirely depends on the way the human interacts with the electronic appliances in a smart home. Similarly, human interaction with household appliances in a smart home can be modeled to produce an efficient household appliance scheduling algorithm. The current research studies on a similar topic exhibit several challenges. These challenges are highlighted as follows:

- The scheduling based on DR requires prior electricity pricing information from the electric supply companies.
- The machine learning algorithm requires a huge amount of household appliances data to be trained. However, human nature is always changing and there is a possibility of suggesting inappropriate scheduling of appliances.
- Integrating Human-Appliances Interaction (HAI) with energy consumption is ignored in previous studies.
- Most of the systems are designed to control the electricity pricing, however, ignoring the energy consumption.
- The scheduling of appliances from a one-time slot of a day to another always maximizes the human's discomfort level.
- Scheduling based on traditional scheduling algorithms are good in a specific scenario, however, most of them result inappropriately in generic smart home scenarios.
- The existing HEMS systems perform inappropriately with a large number of appliances in a smart home.
- Optimizing the working of smart grids does not guarantee the better performance of HEMS systems.

2) CONTRIBUTION

In order to address the aforementioned challenges in the current literature and to efficiently schedule the household appliances, the availability of a smart home user in a particular environment is divided into three different parts using the proposed system based on Q Learning algorithm i.e. 1) A state s is considered as the current situation of an appliance n. For instance, if the appliance is consuming energy it is in ON state otherwise OFF state, 2) An action a is performed by the user i in an environment depending on the category of the appliances as discussed in Section 1, and 3) A reward matrix r with the values representing the worth of an action that will be taken in the next state. After acquiring the information of states, actions, and rewards, each agent connected with appliances computed the Q values for the next action to be taken. The best actions in a state are performed based on the policy function v derived using the RL algorithm. Further, each agent is programmed in a way to consume a limited amount of energy assigned to it. Also, a goal state g is defined as if somehow an agent gets successful in running the appliance in a state within the limits of energy assigned to it. Similarly, a system can have many goal states depending on consuming the energy usage within the limits.

B. AGENT APPLIANCE INTERACTION BASED ENERGY EFFICIENT SMART HOME

1) OVERVIEW

Modeling human behavior, in general, is a complicated job and it requires multiple machine learning models working simultaneously. Also, modeling human behavior in real-time is an impossible task as it requires a lot of time to decide the best action at any generic scenario. However, modeling human behavior in a scenario such as human-computer interaction, human-to-vehicle interaction, etc. can be modeled with the help of a single or multiple machine learning algorithm. Keeping such an assumption in mind, in this research work, we modeled the interaction of humans with electronic appliances using the Q learning reinforcement algorithm. The Q learning model works by incorporating the states, actions, and rewards in a way to choose the most appropriate next action as shown in Figure 5. A particular action can be performed in a state if its reward is high compared to other actions. In this research work, the states, actions, and rewards are defined to model the Q learning problem in a single user smart home scenario. The power rating of the appliance running in a time t is represented with a binary representation. For example, if there are n appliances available in a smart home, their power level can be represented by 2^n levels. A user can perform an action based on the reward matrix. For instance, an appliance can be switched on, if it has a maximum reward in the reward matrix. We assumed that each smart home based on the proposed methodology is assigned with a fixed amount of power level. The agent associate with each smart home appliance as shown in Figure 5 is programmed to turn on, off, and changing the power level of the appliances in a way to consume energy within the limits of power level assigned to the smart home. However, switching off or reducing the power level of an appliance may increase the discomfort level of the smart home user. We, therefore, introduce appliances

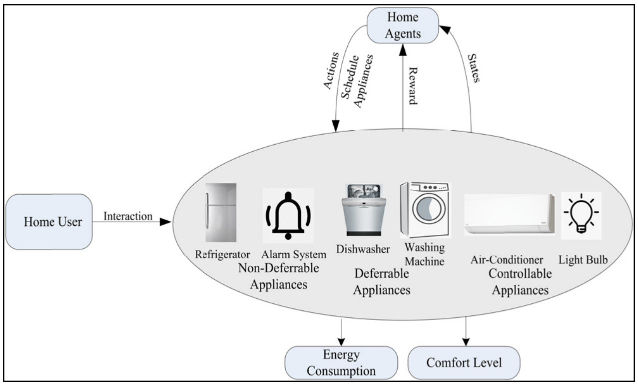

FIGURE 5. The working of the proposed ROSTHA-QL scheme.

priority strategy to turn off or change the power level of those appliances whose priority is low to maximize the comfort level and minimize the energy consumption. The proposed appliances priority strategy turned off an appliance with low priority (an appliance belongs to category C as shown in Section 1) or switches its power level from high to low one at a time. Also, the agents are modeled in a way to find the best sequence of turning off appliances which also ensures turning off the appliances with producing minimum discomfort.

2) WORKING OF PROPOSED RSOTHA-QL SCHEME

This section clearly describes the working of the proposed scheduling scheme for efficient energy management in the smart home scenario following a Q learning model.

The Q learning model is used to optimizes the power consumption of smart homes by modeling the system state, actions, and reward matrix. Further, the agent is programmed with these states, actions, and rewards based on the power ratings and the activity time or operational time of the appliances. The previous activity or the previous human interaction with appliances is used to detect the appropriate and accurate operational time of the appliances using the Q learning. During experiments with the proposed ROSTHA-QL scheme, we observed an appliance that is turned on in the previous time slot is remained on in the next time slot by 56.6∼61.8%. However, the probability of remaining switched on in the next few states reduces linearly and there is a time reached where the appliances are completely switched off or move to an idle state. Also, if a human interacted thrice in a week with an appliance in one specific time slot during the last three days, then we can easily deduce the probability as $P_{ni} = {}^{3}/_{7}$ while considering weekdays and weekends the same. Thus, keeping the above situation in mind, we developed the system state for the agent by incorporating the power rating, load profile, and interaction of the user with appliances. Similarly, a goal state i.e. G is defined by considering a sequence of operating appliances resulting in consuming the energy within the limits assigned to an agent. For instance, let assume an agent is allowed to use a total of energy less or equal to the maximum energy of the smart home represented by $max(E)$. Also, if n appliances are running at a time T then the total energy to reach a G must be less than or equal to $max(E)$ i.e. $\sum_{n=1}^{n} E_n \leq max(E)$. Similarly, in the case of multiple appliances in a smart home, a system can have more than one Gs. However, in each of the G states, an agent is allowed to use an amount of energy which is always less than or equal to $max(E)$. To compute a safe sequence of the appliances running in the smart home, we use a binary coding phenomenon. For instance, a 0 represents an off-state and 1 represents an on-state. Similarly, if there are n number of appliances available there are 2^n power levels available to utilize. For example, if there are 3 appliances available in a smart home, the agent is allowed to use 8 power levels in total.

A state consists of one or more actions. An action can be turned on and off, stop and pause the operation of an appliance, etc. However, in this work, we are interested in three types of operations that are performed by an agent within a state i.e. 1) turning on i.e. T_{ON} and 2) turning off i.e. T_{OFF} an appliance and 3) changing the power level of an appliance. Similarly, an agent is designed to take one action per time. For instance, an agent can switch off an appliance at a time T, however, it cannot switch on another appliance at the same time. Further, the agent moves from one state to another after performing various actions in a state. However, the agent maintains the sequence of actions which leads to G. Also, the agent maintains a system of turning off or changing the power level of an appliance whenever the energy consumption exceeds the total energy E_T. If turning off one appliance does not guarantee a G, then the agent turns off more appliances until and unless the total energy consumption of the smart home environment drops below E_T. The Q learning helps us to find the best sequence of turning off the appliances ensures the appliances with higher priority remained switched on and achieving minimum energy consumption. Also, the sequences guarantee a high comfort level of the smart home user. The agents continuously learn a similar behavior of turning on and off appliances. In this research work, an agent is able to design a policy for switching on or off an appliance based on the operation history of the appliances. For instance, an appliance of high energy consumption was turned on for a longer time in the last week can be restricted from turning on for the longest time in the current week to reduces energy consumption. Thus, designing such a policy requires a historic knowledge of the operation of the appliances last week, month, and even year. However, Q learning is a value iterative reinforcement learning and the agent always generates a new policy at the iteration level which satisfies the Bellman Equation as follows.

$$Q_t'(s_t, a_t) = r(s_t, a_t) + \gamma \times maxQ(s_t, a_t) \tag{1}$$

where $\gamma \in [0, 1]$ represents a discounted factor, which shows the relative importance between the previous and current rewards. For example, if the value of γ is closer or equal to 0, the agents will choose the current rewards otherwise the agent will look for long term rewards. Whenever the agent acts at a time T such that T = [$t_1, t_2, t_3, \cdots, t_n$] the corresponding Q value in the state-action table is updated using the Bellman Equation as follows.

$$Q(s_t, a_t) = Q(s_t, a_t) + \theta\,[r(s_t, a_t) + \gamma \times maxQ(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] \quad (2)$$

where θ is the learning rate, which represents the learning of new Q values to overrides the old Q values. If the value of θ is set to 0 the entire part of the equation on the rightmost side equal to 0, which indicates that the agent learns nothing. However, setting its value to 1 indicates that the agent learns while considering the current estimates and ignoring the prior knowledge. To achieve a balance relation between the new information and the old information, the learning rate needs to be set between 0 and 1. This way, the agents learn new Q values and updated them to the state-action pair in the Q table. Finally, the Q values reach to highest value and thus the convergence of the Q values stops at the highest Q values obtained in the Q table. Interested readers are directed to the reference number [34], [35] for more detailed information.

Also, the Q learning does not rely on old experience or data to help the agents to design policy. To deal with such situations, the appliances' preferences i.e. categorization of the appliances help the agent in designing policy. Before designing the policy, an agent categorizes the home appliances into three types as stated in Section 1. For instance, if an agent noticed the energy consumption is exceeding the limit during an iteration, it switches off or reschedule the appliances belong to category C. Because rescheduling or shifting the appliances to another state or time for the operation is applicable in the case of category C appliances. However, in the middle of an operation, rescheduling or shifting such appliances is not recommending. Also, halting the operation of such appliances in the middle of operation highly increases the discomfort level of a home user. For instance, halting a washing machine in the middle of washing the clothes can increase the discomfort level of the home user. However, the operation of such devices can be shifted to other times of the day before the operating slot. Thus, the devised policy must help in assigning negative rewards to such operations. Also, in such cases, the agent learns using the policy rules to switch off the appliances of category B or A, respectively. Overall, mapping of such behavior of a human to an agent results in achieving high comfort and minimizing the dissatisfaction level. Finally, the optimal policy v i.e. the policy gives high Q value at a time T is achieved during each iteration of the Q learning algorithm as follows:

$$v = \max Q(s_t, a_t) \quad (3)$$

The high Q value represents those conditions where the energy consumption and comfort level are less and high, respectively. Finally, the reward matrix is designed based on turning off an appliance and the amount of discomfort associated with it. An appliance can be switched off until and unless it guarantees less or no discomfort to the user. Thus, the rewards to those actions are set to high which decreases the discomfort level. For instance, assume a user enters the home from an outdoor high-temperature environment. In such situations, the user always prefers to turn on AC instead of an iron box. Thus, the reward of turning on the AC is always high instead of turning on an iron box. However, if turning on an AC exceed the total energy of the smart home, then the agent will switch off or changing the power level of an appliance of low preferences. Thus, we need to design the reward matrix keeping all the possible options to increase the comfort level and decrease the energy consumption. We aim to achieve G based on the maximum Comfort Level (CL) represented by $max(CL)$ of the smart home user and minimum Energy Consumption $min(EC)$ of the appliances. Finally, the reward matric r for all the actions can be computed as follows.

$$r = \begin{cases} 1 & if\ G\ is\ reached\ without\ min\,(CL)\ and\ max(EC) \\ -1 & if\ a\ G\ is\ reached\ with\ min\,(CL)\ and\ max(EC) \\ 0 & Agent\ per\ forms\ nothing \end{cases} \quad (4)$$

The reward for the actions considering the above value of the CL and EC is categorized as 1) 1 – means the highest reward, 2) 0 – nothing, and 3) −1 – means low or negative reward. The values to each operation in the state-action pair is assigned based on the probability of maximizing the comfort level and minimizing the energy consumption. Also, each home appliance is assigned a dissatisfaction value based on the level of dissatisfaction caused due to switching off an appliance in each category using the following relation [36].

$$\varphi_n = e^{\beta_n\left(1-\left(\frac{E_n}{m_n}\right)\right)} - 1, \quad \beta_n > 0 \quad (5)$$

where m_n represents the median level of energy demand represented with E. The energy demand is decreasing continuously and upon reaching the median level the value of φ_n changes from positive to negative. Where β_n represent the priority factor of an appliance. For instance, if an appliance belongs to a lower category C or β_n is high, the demand for E is low and vice versa. Therefore, in such cases where an appliance of low priority is switched off resulting in lowering the dissatisfaction level. Also, in the cases of deferrable devices that belong to category C, turning off appliances results in increasing the discomfort level for a smaller amount. In most cases, turning off appliances of such a category does not affect the overall performance of the system.

Finally, based on the above discussion, an agent is programmed to achieve two objectives one to minimize the energy consumption if it is exceeding the energy limit i.e. max(E) and secondly to minimize the dissatisfaction level by using Equation 5. Following Algorithm 2 is used to model the entire proposed scheme to achieve the above two objectives.

Algorithm 2 ROSTHA-QL Based Scheduling of Operational Time of Home Appliances

1 **Input:** List of actions, rewards, and states with policies and appliances priorities
2 **Output:** Optimal Turning ON and OFF appliances with minimum discomfort and energy consumption
3 appliances [ ] $\leftarrow$ PR
4 initialize maximum energy of the smart home
5 initialize β values
6 $\gamma \leftarrow 0.9$
7 actions a [ ] $\leftarrow T_{ON}, T_{OFF}$
8 for each agent **do**
9 initialize Q-value to **0**
10 initialize θ to **0.1**
11 for each iteration **Repeat**
12 initialize s_t
13 for each t **Repeat**
14 select a_t from s_t using the **ϵ-greedy policy**
15 for $\forall$ $\boldsymbol{a_t}$
16 if a_t $\leftarrow$ switching off && Appliance $\in$ A
17 do not select a_t
18 goto 14
19 else
20 select a_t
21 perform an action a_t and observe $r\left(s_t, a_t\right)$ and next s_{t+1}
22 $Q\left(s_t, a_t\right) = Q\left(s_t, a_t\right) + \theta\left[r\left(s_t, a_t\right) + \gamma \times maxQ\left(s_{t+1}, a_{t+1}\right) - Q\left(s_t, a_t\right)\right]$
23 check if s_{t+1} is a terminal state
24 check if Q-values are converged
25 output the optimal appliances scheduling at time $\boldsymbol{t}$
26 end for
27 end for

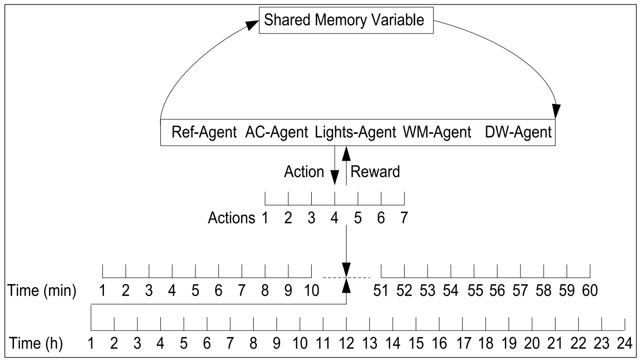

FIGURE 6. Simulation environment.

TABLE 1. Parameters used in the proposed model.

Device Type	βi	Power Rating (W)	Operation Period	Remarks
Ref	2.0	0 ~ 500	24 h	Everyday
AC	2.5	0 ~ 500	12 h	Everyday
L1	3.0	200 ~ 500	18:00 PM ~ 5:00 AM	Everyday
L2	3.0	200 ~ 500	18:00 PM ~ 12:00 AM	Everyday
WM	3.5	800 ~ 1000	10:00 AM ~ 12:00 PM OR 20:00 ~ 22:00 PM	Once a week
DW	4.0	600 ~ 800	10:00 AM ~ 12:00 PM & 20:00 ~ 22:00 PM	Everyday

V. PERFORMANCE EVALUATION

A. SIMULATION SETUP

In this section, an extensive simulation is performed to test the working of the proposed scheme for 24 hours a day. Each hour is further divided into 60 minutes as shown in Figure 6. The agents are programmed to perform an action from 7 different actions at a time in a minute of an hour. The action is then compared with the shared memory variable holding the total energy currently consuming in the smart home. Thus, an action is always taken whenever it does not exceed the maximum energy consumption threshold. Further, the simulation parameters are shown in Table 1. These parameters are mostly derived from [36]–[38]. However, these values depend on the nature of the appliances and production companies. Similarly, the power ratings are also different depending on the price of the appliances, however, it does not affect the simulation environment and the working of the proposed scheme. To explore all the states-action pairs, initially, the ϵ-greedy policy is set to as minimum as 0.1. Similarly, the learning rate θ and the discounted factor γ is set to 0.1 and 0.9, respectively. The maximum energy limit in a state is set to 1500 watts. However, to maintain coordination between the states and ensure the energy should not exceed the maximum energy limit, the consumption of the energy in a previous state is shared with the current state. Further, the system used for the simulation is comprised of Intel(R) Core i5 CPU with a clocking rate of 3.10 GHz, a RAM of 8 GB and Python 3.7 in Spyder 3.3.6.

To be realistic, the peak, semi peak, and off-peak times are considered for efficient utilization of time slots during the entire day as shown in Table 2. In other words, the categories of appliances such as B and C are mostly scheduled in the off-peak times of the day. Thus, this well reduces energy consumption as well as energy prices. However, optimizing the prices of the energy is out of the scope of this paper.

Further, a smart home scenario is considered with six appliances i.e. refrigerator (Ref), air-conditioning (AC) system, light bulbs (L1 and L2), washing machine (WM), and dishwasher (DW). The Ref belongs to category A i.e.

TABLE 2. Day time division in peak, semi-peak and off-peak slots.

Time Division	Hours
Peak	1:00 p.m. – 8:00 p.m.
Semi Peak	6:00 a.m. – 1:00 p.m. & 8:00 p.m. – 12:00 a.m.
Off-Peak	12:00 a.m. – 6:00 a.m.

non-deferrable category, the AC, L1, and L2 belongs to category B i.e. controllable category, and WM, and DW belongs to category C i.e. deferrable category. The comfort values are directly proportional to the category of the appliances. Similarly, each agent attached to an appliance first check the total energy of the home and then act as switching on an appliance. Thus, in the simulation, all the agents communicated with each other in a coordinated environment using a shared-memory variable holding the total energy at a current moment. Thus, an agent performs an action based on the reward of the action plus the current state of the other appliances i.e. total energy consumption of the home.

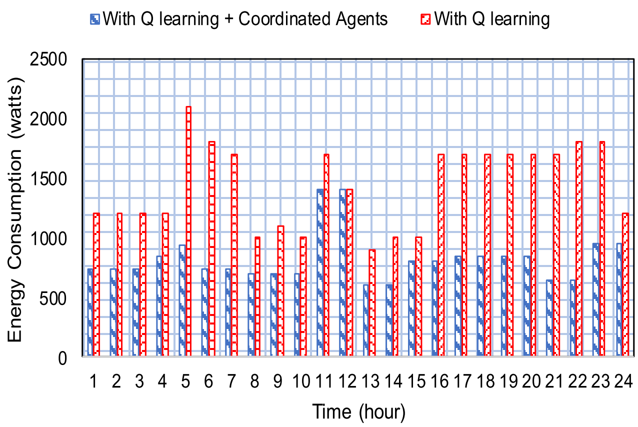

FIGURE 7. **Energy consumption with and without Q learning and coordinated agents.**

B. RESULTS AND DISCUSSIONS

If an agent is switching on an appliance, then those appliances with lower priority can be switched off or changes to a lower power level to ensure that the power consumption will not exceed the limit. This coordination helps in reducing the energy consumption of the entire home. The energy consumption compared to an environment with and without Q learning and coordinated agents are shown in Figure 7. Also, Figure 7 shows that as the energy consumption starts exceeding the limits by turning on the appliances, the agents attached to the low priority appliances are triggered to schedule back to another power level or switched off at the moment to reduce the energy consumption.

Similarly, in Figure 8, the proposed scheme is compared with scheduling based on the DR strategy. As we know, the scheduling based on DR always fulfills the demand of the consumer from the smart grid. Therefore, the smart grid does not care about the overall energy consumption of the smart home. For instance, a home user can switch on as many appliances until and unless its demand is fulfilled. Thus, such systems do not incorporate the human interaction for controlling the overall energy consumption of the smart home appliances autonomously. Thus, we can see in Figure 8, the energy consumption of the scheduling based on DR is high on t = 3-4 and 6-20.

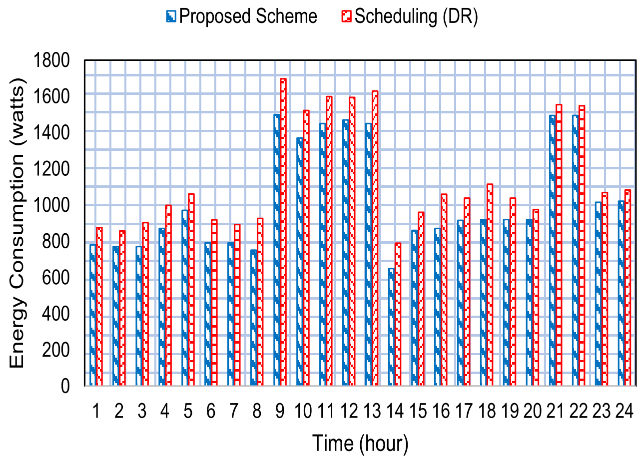

FIGURE 8. **Energy consumption of the proposed scheme against scheduling based on DR.**

The energy consumption of various household appliances is shown in Figure 9. The proposed scheme shows a significant improvement in reducing the energy consumption due to coordination of the agents and scheduling the appliances while considering the user comfort level. Also, Figure 9 shows that as other appliances of high priority are switched on the appliances with low priority are switched to a minimum power level or schedule to next time slots.

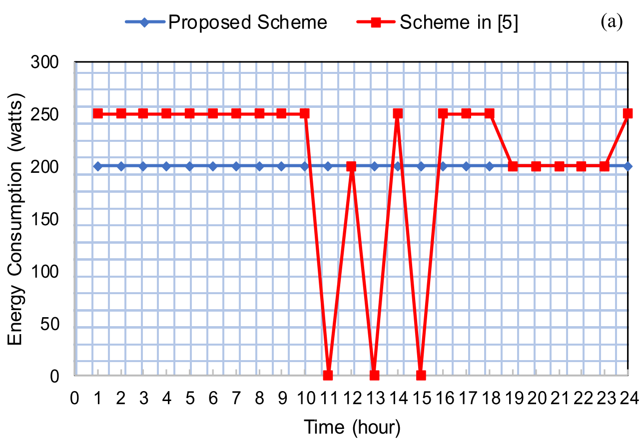

FIGURE 9. **Energy consumption of various household appliances, a) Refrigerator, b) Air conditioner, c) Lights and d) washing machine.**

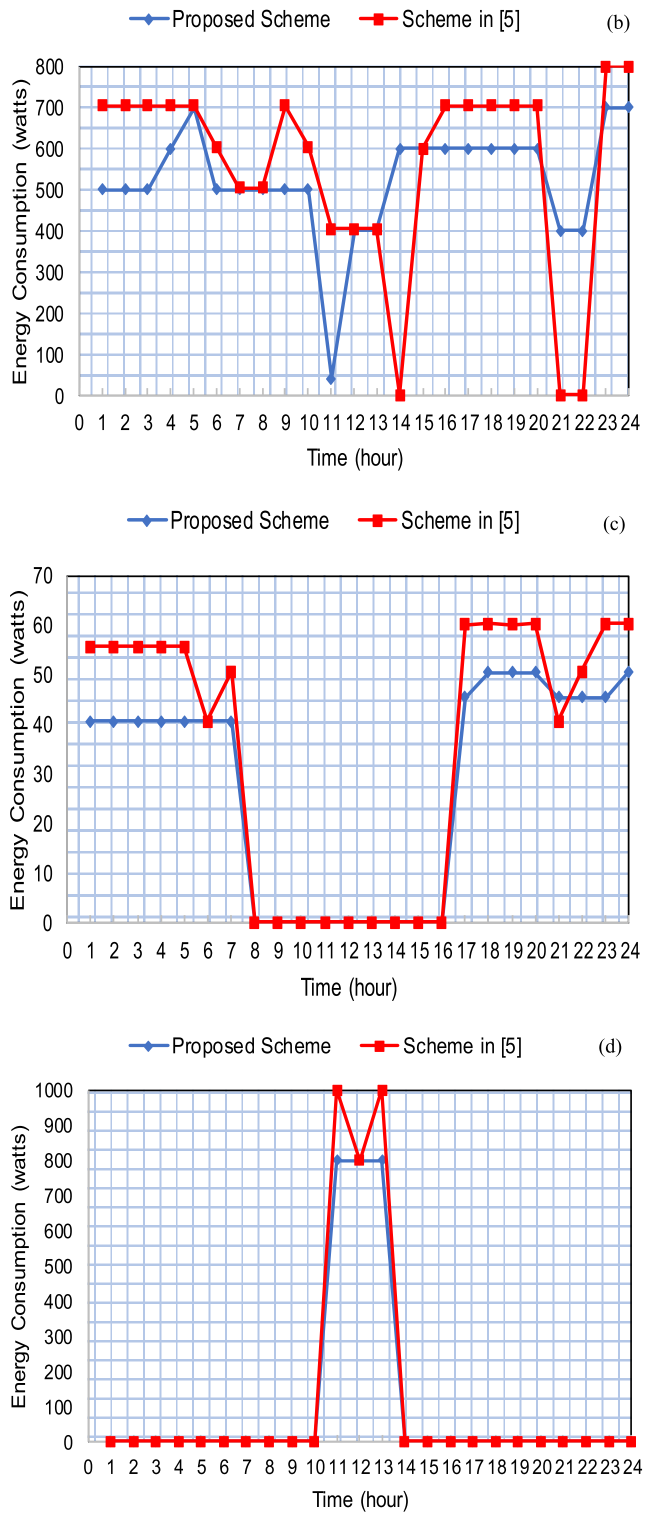

FIGURE 9. ***(Continued.)*** **Energy consumption of various household appliances, a) Refrigerator, b) Air conditioner, c) Lights and d) washing machine.**

We can see in the case of the proposed scheme in Figure 9 (b), the air conditioner is switched to minimum power level at time t = 11, due to the switching on of another appliance at t = 11. Similarly, the energy consumption of air conditioner is dropped to 0 watts in the case of LST based scheduling at t = 14, 21, and 22 because switching on the AC at these specific times can exceed the maximum energy consumption limit i.e. 1500 watts. Similarly, the energy consumption of refrigerator as shown in Figure 9 (a) is dropped to a minimum level of 0 watts at time t = 11, 13, and 15 in the case of LST based scheduling due to exceeding the maximum energy limit. However, in the case of the proposed scheme, the refrigerator is always in on-state because it belongs to category A appliances. Thus, such scheduling of appliances based on the proposed scheme reduces the energy consumption of the appliances, however, keeping the dissatisfaction level of the user at a minimum level. Also, the proposed scheme keeping track of all those appliances which belong to category A. Because switching off those appliances highly increases the discomfort level of the user.

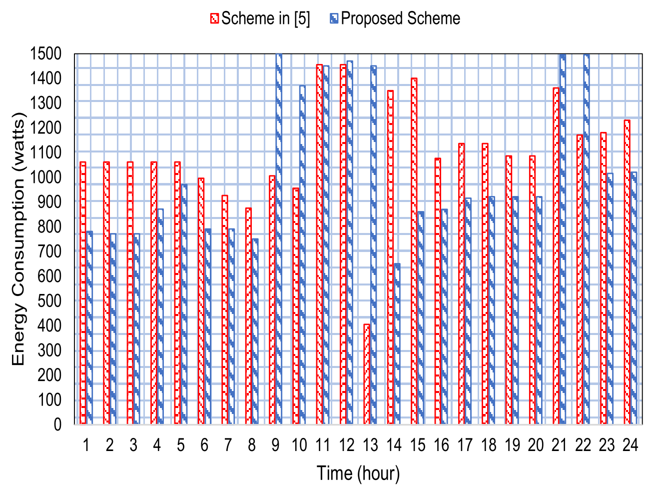

FIGURE 10. Energy consumption of the proposed scheme throughout the day.

Figure 10 shows the energy consumption of the proposed system with our previous study based on the LST scheduling algorithm. During the experimental results, we set the maximum energy threshold to 1500 watt per hour. For instance, if the energy consumption of the appliances exceeds this threshold, the proposed scheme changes the power level of an appliance or in the worst case it switches off an appliance to meet the optimal energy consumption i.e. less than 1500 watts. On the other hand, the LST performs a similar approach, however, the LST mechanism only switches off the appliance to meet the energy consumption limit. Switching off an appliance increases the discomfort level of home user and it is, therefore, reveals from the experiment that switching off an appliance or do not let an appliance turn on is not an appropriate decision. Also, the LST does not consider the priority level of the appliances and thus switches off any appliance even the refrigerator. The dishwasher is scheduled to turn on at 8:00∼9:00 AM ($t = 9 \sim 10$), however, turning on the dishwasher will increase the consumption of both the proposed scheme and LST based scheduling algorithm from 1500 watts. In the case of the proposed scheme, the dishwasher is turned on as scheduled, however, the power level of the AC is reduced from 500 watts to 400 watts to keep the energy consumption within the limit i.e. 1500 watts. On the other hand, the LST based scheduling does not allow the turning on of the dishwasher at 8:00∼9:00 AM.

The activity of the dishwasher is rescheduled to 13:00∼14:00 PM (t = 14 ∼ 15) of the same day, however, the AC is switched off to set the energy consumption within the limits. Also, we noticed that the rescheduling of the dishwasher increases the dissatisfaction of the home user as shown in Figure 11. As the temperature is normally hot during 13:00 and 14:00 PM, therefore, turning off AC at this time may results in inappropriate scheduling. Also, 13:00 and 14:00 PM is considered as peak hours thus highly increases the electricity tariff at these times. The experiment also reveals that the LST and similar scheduling algorithms do not consider the satisfaction of the home user in scheduling the appliances. Even in the case of DR-based scheduling,

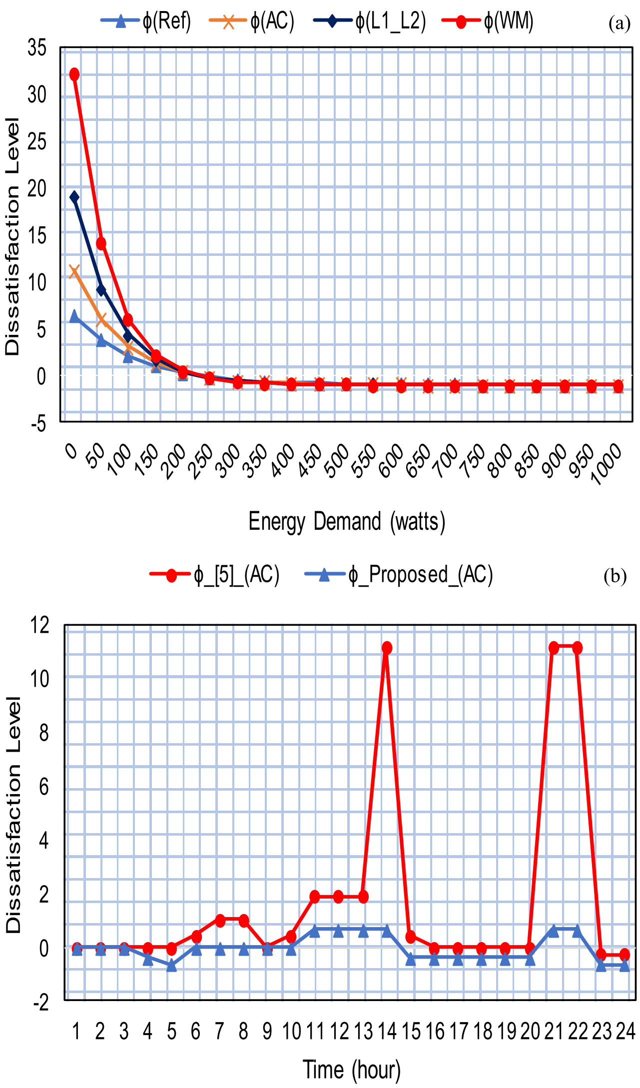

FIGURE 11. Analysis of user dissatisfaction level in the case of a) generic data from 0 to 1000 watts, b) energy data from the operation of AC in 24 hours, c) energy data from the operation of the refrigerator in 24 hours d) energy data from the operation of Lights in 12 hours.

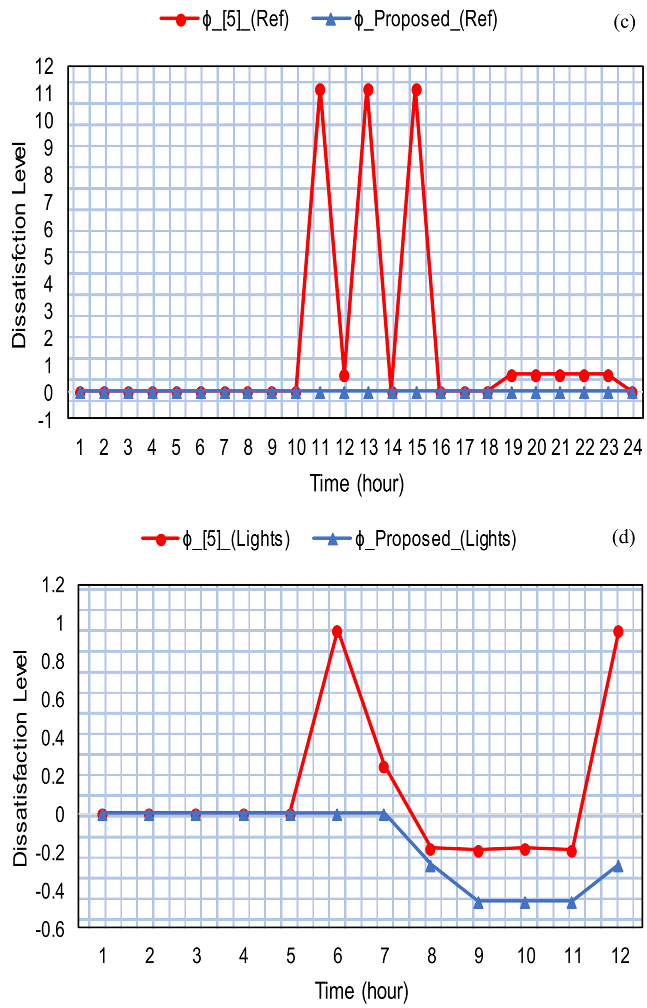

FIGURE 11. *(Continued.)* Analysis of user dissatisfaction level in the case of a) generic data from 0 to 1000 watts, b) energy data from the operation of AC in 24 hours, c) energy data from the operation of the refrigerator in 24 hours d) energy data from the operation of Lights in 12 hours.

the satisfaction of the home user is neglected or avoided when turning on or off an appliance to meet the optimal energy consumption [39]. Thus, learning-based models such as Q learning in the case of the proposed scheme always consider user satisfaction in optimizing energy consumption via scheduling.

Figure 11 shows the results of the dissatisfaction level of the user while using the proposed scheme and the scheme in [5]. The scheme in [5] does not concatenate the comfort level of the user with the appliances. The dissatisfaction level of the user against the energy demand is shown in Figure 11 (a) considering energy values range from 0 to 1000 watts. As we go down the curve the energy demand increases, and the dissatisfaction level decreases. The increase in energy demands fulfills the need of the home user and thus the user dissatisfaction level decreases. However, if the energy demands do not meet, the user dissatisfaction level increases. Thus, whenever, an appliance of high priority such as refrigerator or AC is switched off resulting in an increase of dissatisfaction to the user as shown in Figure 11 (b). On the other hand, the proposed scheme never turns off those appliances whose priority is high. In other words, the reward of turning off those appliances is set to minimum or negative. Thus, during the training phase, the agents always learn to turn off the appliances with high priority results in less reward. Figure 11 (b) shows that the dissatisfaction level at t = 11-14 and 22-23 in the case of LST based scheduling is significantly high. The reason for such an increase in dissatisfaction level at t = 11-14 and 22-23 is that the LST based scheduling switched off the AC with high priority. Also, the LST based scheduling does not follow an intelligent model and thus always performing inappropriate in learning the environment. The proposed scheme always learns the environment and updated the Q values based on the experience from the environment. Therefore, the proposed scheme always performs an action that decreases energy consumption and discomfort.

VI. CONCLUSION

In this paper, we modeled the human appliance interaction using reinforcement learning to achieve minimum energy consumption and lower discomfort of the home users in a smart home scenario. The well-known learning algorithm called the Q learning is used to train the agents attached with each household appliance to perform actions of turning on, off, and changing power levels according to the maximum energy consumption constraint. The entire day is further divided into twenty-four-hour slots to schedule the household appliances in each hour. The time slots are further categorized into three groups i.e. 1) peak, 2) semi-peak, and 3) off-peak to efficiently schedule the deferrable, non-deferrable, and controllable appliances. The actions performed by an agent is communicated with the rest of the agents to establish agent coordination with each other. The performance analysis showed that the proposed scheme scheduled the household appliances with minimum energy consumption and minimum discomfort of the smart home user. The simulation results are compared with our previous study based on scheduling the household appliances using the LST scheduling algorithm. A significant improvement is shown comparing the proposed scheme with the LST based scheduling in the context of reducing energy consumption and minimizing the dissatisfaction level of the home user.

In future work, a multiple home user scenario will be modeled with the help of a deep Q learning algorithm and artificial neural networks. The agents of a community smart homes will be integrated to schedule the operational time of the household appliance of the entire smart community. Also, an edge computing paradigm will be adopted to control the entire communication of the home from the edge. Resulting in a distributed environment for future DR systems.

REFERENCES

[1] B. Zhou, W. Li, K. W. Chan, Y. Cao, Y. Kuang, X. Liu, and X. Wang, "Smart home energy management systems: Concept, configurations, and scheduling strategies," *Renew. Sustain. Energy Rev.*, vol. 61, pp. 30–40, Aug. 2016.

[2] B. Celik, R. Roche, S. Suryanarayanan, D. Bouquain, and A. Miraoui, "Electric energy management in residential areas through coordination of multiple smart homes," *Renew. Sustain. Energy Rev.*, vol. 80, pp. 260–275, Dec. 2017.

[3] I. Abubakar, S. N. Khalid, M. W. Mustafa, H. Shareef, and M. Mustapha, "Application of load monitoring in appliances' energy management—A review," *Renew. Sustain. Energy Rev.*, vol. 67, pp. 235–245, Jan. 2017.

[4] H.-T. Roh and J.-W. Lee, "Residential demand response scheduling with multiclass appliances in the smart grid," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 94–104, Jan. 2016.

[5] B. Silva, M. Khan, and K. Han, "Load balancing integrated least slack time-based appliance scheduling for smart home energy management," *Sensors*, vol. 18, no. 3, p. 685, Feb. 2018.

[6] A. Mukherjee, P. Mukherjee, N. Dey, D. De, and B. K. Panigrahi, "Lightweight sustainable intelligent load forecasting platform for smart grid applications," *Sustain. Comput., Informat. Syst.*, vol. 25, pp. 1–18, 2020.

[7] F. Pallonetto, M. De Rosa, F. Milano, and D. P. Finn, "Demand response algorithms for smart-grid ready residential buildings using machine learning models," *Appl. Energy*, vol. 239, pp. 1265–1282, Apr. 2019.

[8] R. Lu and S. H. Hong, "Incentive-based demand response for smart grid with reinforcement learning and deep neural network," *Appl. Energy*, vol. 236, pp. 937–949, Feb. 2019.

[9] K. Chauhan, S. Jani, D. Thakkar, R. Dave, J. Bhatia, S. Tanwar, and M. S. Obaidat, "Automated machine learning: The new wave of machine learning," in *Proc. 2nd Int. Conf. Innov. Mech. Ind. Appl. (ICIMIA)*, Bangalore, India, Mar. 2020, pp. 205–212.

[10] G. Le Ray and P. Pinson, "The ethical smart grid: Enabling a fruitful and long-lasting relationship between utilities and customers," *Energy Policy*, vol. 140, pp. 1–12, 2020.

[11] S. Tanwar, P. Patel, K. Patel, S. Tyagi, N. Kumar, and M. S. Obaidat, "An advanced Internet of Thing based security alert system for smart home," in *Proc. Int. Conf. Comput., Inf. Telecommun. Syst. (CITS)*, Dalian, China, 2017, pp. 25–29.

[12] B. N. Silva, M. Khan, and K. Han, "Internet of Things: A comprehensive review of enabling technologies, architecture, and challenges," *IETE Tech. Rev.*, vol. 35, no. 2, pp. 205–220, Mar. 2018.

[13] M. Khan, K. Han, and S. Karthik, "Designing smart control systems based on Internet of Things and big data analytics," *Wireless Pers. Commun.*, vol. 99, no. 4, pp. 1683–1697, Apr. 2018.

[14] T. S. Thakkar, "Combining user-based and item-based collaborative filtering using machine learning," in *Information and Communication Technology for Intelligent Systems*. Singapore: Springer, 2019.

[15] F. A. Qayyum, M. Naeem, A. S. Khwaja, A. Anpalagan, L. Guan, and B. Venkatesh, "Appliance scheduling optimization in smart home networks," *IEEE Access*, vol. 3, pp. 2176–2190, 2015.

[16] X. Hou, J. Wang, T. Huang, T. Wang, and P. Wang, "Smart home energy management optimization method considering energy storage and electric vehicle," *IEEE Access*, vol. 7, pp. 144010–144020, 2019.

[17] D. M. Minhas and G. Frey, "Modeling and optimizing energy supply and demand in home area power network (HAPN)," *IEEE ACCESS*, vol. 8, pp. 2052–2072, 2019.

[18] S. Bahrami and A. Sheikhi, "From demand response in smart grid toward integrated demand response in smart energy hub," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 650–658, Mar. 2016.

[19] M. S. Munir, S. F. Abedin, N. H. Tran, and C. S. Hong, "When edge computing meets microgrid: A deep reinforcement learning approach," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7360–7374, Oct. 2019.

[20] T. Singh. *Smart Home Dataset with weather Information*. Accessed: May 26, 2020. [Online]. Available: https://www.kaggle.com/taranvee/smart-home-dataset-with-weather-information

[21] M. Pipattanasomporn, M. Kuzlu, S. Rahman, and Y. Teklu, "Load profiles of selected major household appliances and their demand response opportunities," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 742–750, Mar. 2014.

[22] L. Liu, X. Yang, H. Huang, and S. Hu, "Smart home scheduling for cost reduction and its implementation on FPGA," *Smart Home Scheduling Cost Reduction Implement.*, vol. 24, no. 4, pp. 1–15, 2015.

[23] A. Alirezazadeh, M. Rashidinejad, A. Abdollahi, P. Afzali, and A. Bakhshai, "A new flexible model for generation scheduling in a smart grid," *Energy*, vol. 191, pp. 1–17, Oct. 2020.

[24] L. Liu, Y. Liu, L. Wang, A. Zomaya, and S. Hu, "Economical and balanced energy usage in the smart home infrastructure: A tutorial and new results," *IEEE Trans. Emerg. Topics Comput.*, vol. 3, no. 4, pp. 556–570, Dec. 2015.

[25] N. G. Paterakis, O. Erdinc, A. G. Bakirtzis, and J. P. S. Catalao, ''Optimal household appliances scheduling under day-ahead pricing and load-shaping demand response strategies,'' *IEEE Trans. Ind. Informat.*, vol. 11, no. 6, pp. 1509–1519, Dec. 2015.
[26] H. Wang, T. Huang, X. Liao, H. Abu-Rub, and G. Chen, ''Reinforcement learning in energy trading game among smart microgrids,'' *IEEE Trans. Ind. Electron.*, vol. 63, no. 8, pp. 5109–5119, Aug. 2016.
[27] M. Rahmani-Andebili, ''Scheduling deferrable appliances and energy resources of a smart home applying multi-time scale stochastic model predictive control,'' *Sustain. Cities Soc.*, vol. 32, pp. 338–347, Jul. 2017.
[28] K. Murugaperumal and P. A. D. V. Raj, ''Integrated energy management system employing pre-emptive priority based load scheduling (PEPLS) approach at residential premises,'' *Energy*, vol. 186, pp. 1–12, Jan. 2019.
[29] A. Jindal, B. S. Bhambhu, M. Singh, N. Kumar, and K. Naik, ''A heuristic-based appliance scheduling scheme for smart homes,'' *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3242–3255, May 2020.
[30] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, ''Human-level control through deep reinforcement learning,'' *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.
[31] D. Gariba and B. Pipaliya, ''Modelling human behaviour in smart home energy management systems via machine learning techniques,'' in *Proc. Int. Autom. Control Conf. (CACS)*, Taichung, Taiwan, Nov. 2016, pp. 53–58.
[32] J. J. C. Magdalene and S. E. Zoraida, ''Predicting household energy consumption in smart grid based on seasonality using stochastic Markov chain,'' *J. Study Res.*, vol. 12, no. 2, pp. 93–100, 2020.
[33] F. Ruelens, B. J. Claessens, S. Quaiyum, B. De Schutter, R. Babuska, and R. Belmans, ''Reinforcement learning applied to an electric water heater: From theory to practice,'' *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3792–3800, Jul. 2018.
[34] T. Jaakkola, M. I. Jordan, and S. P. Singh, ''On the convergence of stochastic iterative dynamic programming algorithms,'' *Neural Comput.*, vol. 6, no. 6, pp. 1185–1201, Nov. 1994.
[35] F. S. Melo, ''Convergence of Q-learning: A simple proof,'' Inst. Syst. Robot., Lisbon, Portugal, Tech. Rep., 2001, pp. 1–4.
[36] M. Yu and S. Ho Hong, ''A real-time demand-response algorithm for smart grids: A stackelberg game approach,'' *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 879–888, Mar. 2016.
[37] X. H. Li and S. H. Hong, ''User-expected price-based demand response algorithm for a home-to-grid system,'' *Energy*, vol. 64, pp. 437–449, Jan. 2014.
[38] K. Ma, T. Yao, J. Yang, and X. Guan, ''Residential power scheduling for demand response in smart grid.,'' *Int. J. Electr. Power Energy Syst.*, vol. 78, pp. 320–325, 2016.
[39] F. Alfaverh, M. Denai, and Y. Sun, ''Demand response strategy based on reinforcement learning and fuzzy reasoning for home energy management,'' *IEEE Access*, vol. 8, pp. 39310–39321, 2020.

MURAD KHAN received the B.S. degree in computer science from the University of Peshawar, Pakistan, in 2008, and the master's and Ph.D. degrees in computer science and engineering from the School of Computer Science and Engineering, Kyungpook National University, Daegu, South Korea. He has published over 80 international conference and journal papers along with two books chapters and edited books in Springer and CRC press. His research interests include ad-hoc and wireless networks, architecture designing for the Internet of Things, and communication protocols designing for smart cities and homes, big data analytics, and so on.

JUNHO SEO received the M.S. degree from Kyungpook National University, Daegu, South Korea, where he is currently pursuing the Ph.D. degree in computer science and engineering. He is a member of the Wireless and Mobile Internet Laboratory (MoNet). His research interests include the Internet of Things, the Internet of Underwater Things, adaptation layers, such as constrained application protocol, and IPv6 over low-power wireless personal area networks.

DONGKYUN KIM (Member, IEEE) received the B.S. degree from Kyungpook National University and the M.S. and Ph.D. degrees from Seoul National University, South Korea. He was a Visiting Researcher with the Georgia Institute of Technology, Atlanta, GA, USA. He also performed a Postdoctoral Program at the University of California at Santa Cruz. He is currently a Professor with the School of Computer Science and Engineering, Kyungpook National University, Daegu. His research interests include ad-hoc networks, sensor networks, and wireless LAN. He has been a TPC Member of several IEEE conferences. He received the Best Paper Award from the Korean Federation of Science and Technology Societies, in 2002.