Residential Load Scheduling With Renewable Generation in the Smart Grid: A Reinforcement Learning Approach

Remani T, E. A. Jasmin, and T. P. Imthias Ahamed

***Abstract*—The significance and need of demand response (DR) programs is realized by the utility as a means to reduce the additional production cost imposed by the accelerating energy demand. With the development in smart information and communication systems, the price-based DR programs can be effectively utilized for controlling the loads of smart residential buildings. Nowadays, the use of stochastic renewable energy sources like photovoltaic (PV) by a small domestic consumer is increasing. In this paper, a generalized model for the residential load scheduling or load commitment problem (LCP) in the presence of renewable sources for any type of tariff is presented. Reinforcement learning (RL) is an efficient tool that has been used to solve the decision making problem under uncertainty. An RL-based approach to solve the LCP is also proposed. The novelty of this paper lies in the introduction of a comprehensive model with implementable solution considering consumer comfort, stochastic renewable power, and tariff. Simulation experiments are conducted to test the efficacy and scalability of the proposed algorithm. The performance of the algorithm is investigated by considering a domestic consumer with schedulable and nonschedulable appliances along with a PV source. Guidelines are given for choosing the parameters of the load.**

***Index Terms*—Demand response (DR), distributed generation (DG), load scheduling, photovoltaic (PV) source, reinforcement learning (RL), smart grid.**

Nomenclature

η	Learning parameter.
γ	Discount factor.
ϵ	Exploration parameter.
α, β	Parameters of the Beta distribution function.
a_k	Action taken in the k^{th} time slot.
a_g	Greedy action.
$j \in \{1 \ldots m\}$	Index denoting the load.
l_j	ON duration of j^{th} load.
r_j	Power rating of the load in kilowatt.
u_j^k	Status of j^{th} load during k^{th} time slot.
u_{ji}^k	Status of j^{th} load during k^{th} time slot with i^{th} source.
$O(a)$	Reward for action a.
$q(a)$	Expected value of the reward, $O(a)$.
x_k	$[x_1(k), x_2(k)]$ state of the system at k^{th} slot.
$g(x, a, x_{\text{new}})$	Cost incurred when the process moves from x to x_{new}.
$\mathcal{A}$	Action set.
χ	State space.
C^k	Energy cost per kWh for the k^{th} time slot.
$Q(x, a)$	Q value of state action pair (x, a).
$Q^n(x, a)$	Estimated Q value of state action pair (x, a) at n^{th} iteration.
$Q^*(x, a)$	Optimal Q value of state action pair (x, a).
P_S	Power output of the PV module.
V_{oc}	Open-circuit voltage in volt.
I_{sc}	Short-circuit current in ampere.
V_{MPP}	Voltage at maximum power point in volt.
I_{MPP}	Current at maximum power point in ampere.
T_c	Cell temperature in degree Celsius.
T_A	Ambient temperature in degree Celsius.
K_v	Voltage temperature coefficient V/° C.
K_i	Current temperature coefficient I/° C.
N_{OT}	Nominal operating temperature of cell in degree Celsius.
s	Solar irradiance in kW/m^2.
FF	Fill factor.

I. Introduction

ONE of the major operational issues in the power system operation is balancing generation and load. In the past, big utilities were supplying power to big and small loads. The balancing of generation and loads was achieved by scheduling the generating units. These problems were well formulated as unit commitment problem (UCP), economic dispatch (ED), and automatic generation control and different solutions were proposed and the area has matured [1]. In 1980s, for example, [2], it was realized that the balancing act could be supported by big consumers. It is well known that there is a large-scale deployment of small renewable resources within the consumer premises [3]. With the development in power electronic interface and decrease in the cost of photovoltaic (PV) panels [4], [5], this trend is expected to grow. With the development in communication and automation technologies, many consumer loads have become controllable. Utilities have realized that through various

Manuscript received February 15, 2018; revised April 19, 2018; accepted June 23, 2018. *(Corresponding author: Remani T.)*

Remani T. and E.A. Jasmin are with the Department of Electrical and Electronics Engineering, Government Engineering College, Thrissur 680009, India (e-mail: remaniashok@gmail.com; eajasmin@gectcr.ac.in).

T. P. I. Ahamed is with the Thangal Kunju Musaliar College of Engineering, Kollam 691005, India (e-mail: imthiasa@gmail.com).

Digital Object Identifier 10.1109/JSYST.2018.2855689

demand response (DR) programs under the smart grid paradigm, cooperation of consumers can be utilized in an efficient way for the balancing act [6], [7]. That is, it is possible to bring down the cost of production by shifting the load rather than starting expensive generators. To achieve the aforementioned goal, there are various DR programs. One major thread is the price-based DR (PBDR) program. The aim of the PBDR program is to design a tariff to motivate the consumer to reshape his load curve so that the maximum demand on the system is reduced. The success of the PBDR program involves the following.

1) Design of a tariff by the utility or the distribution company.
2) Scheduling of the loads by consumer.

Utilities have designed various tariffs such as time of use pricing (ToUP), critical peak pricing (CPP), and real-time pricing (RTP). In general, consumer will not try to manually schedule the loads. Thus, there is a need for scheduling algorithms and automation devices.

There is a large volume of work proposing various algorithms and models for the residential DR. Mohsenian *et al.* proposed an incentive-based energy consumption scheduling scheme in which a common energy resource is shared by several residential customers [8]. Each customer is equipped with an automatic energy consumption scheduler. The game theory is used to formulate an energy consumption scheduling game among the users. The objective is to minimize the peak-to-average ratio and the energy cost. A quadratic energy cost function is assumed with two different coefficients. A survey on DR programs is presented in [9], which gives an insight into the qualitative nature of this problem. Muratori *et al.* proposed a dynamic energy management framework for the residential DR by scheduling controllable loads including plug-in electric vehicles [10]. The optimal schedule corresponding to the minimum cost function is obtained using the dynamic programming. The consumer electricity-related expenditures is chosen as the cost function and a multitime-varying electricity pricing scheme is also proposed to eliminate the rebound peak. A detailed review of residential DR programs is presented in [11]. However, research in this area is still in its early stages. Many researches assume that the consumer will try to reduce the maximum demand [12], [13]. Some other researches try to combine the aforementioned two issues, via design of tariff and framing of policies by utilities and scheduling of consumer loads [14], [15]. It is evident that at least initially for the PBDR to succeed, formulation of policies and tariff by utilities and scheduling of loads should be seen independently. If it has to be assumed that consumers will cooperate among themselves or with the utility, policies have to be formulated. So, it is prudent to work either in framing policies to motivate consumers studying the consumer behavior or to come up with scheduling algorithms for a given tariff. The development in the scheduling of generation sources in the past is partially due to the generalized mathematical model for the UCP and ED. We see a lot of parallelism between the UCP and load scheduling, so we call the load scheduling problem as the load commitment problem (LCP).

The major aspects to be considered with respect to the residential load scheduling include comfort level associated with different appliances, constraints on operation of devices, availability, price, and nature of energy supply. This paper formulates the scheduling problem addressing these factors from a consumer perspective. Even though there are many research works [16]–[21] proposing various models from a consumer perspective, there is a lot of divergence in the terminology and models used. For example, loads are classified and termed as elastic and inelastic in [16] and as flexible and inflexible in [18] and [19]. Loads are also categorized as noninterruptible and nonschedulable, interruptible and nonschedulable, and schedulable [20]. Atomic and nonatomic loads are termed as nonpreemptive and preemptive shedulable loads, respectively, in [15] and [20]. In the proposed model, we have introduced a unified terminology as critical loads for all inflexible/nonshedulable/inelastic/baseline loads. Also flexible/shedulable loads with nonpreemptive and preemptive status are termed as atomic and nonatomic, respectively. Based on the class of loads, the start time, finish time, and the duration for which the load has to be ON are assumed. The nomenclature has not yet converged. The proposed work will also be a first step in developing a generalized model for the LCP.

As the cost of electrical energy will be only a fraction of the total household budget, consumer will be ready to schedule their load only if it does not affect their convenience. This is not addressed in majority of the works reported [16], [17], [20]. Some of the research in this direction are presented in [18] and [21], considering inconvenience caused to the consumer also. However, in [18], same inconvenience cost for all loads is assumed. A multiobjective optimization problem where one of the objective is to maximize the user convenience level is reported in [21]. Priority levels are assigned for different appliances. But a quantitative modeling of household devices based on the comfort level still lacks in these works. A parameter udc incorporated in the proposed load model serves to capture the degree of discomfort [22].

Load management for the domestic energy system with the distributed generation (DG) source under the RTP is proposed in [17]. The load management approach includes DG-based scheduling and RTP-based scheduling. The development of a control system for the residential sector with DG is described in [23]. It incorporates the local PV energy generation and the controller maximizes self-consumption. The heuristic algorithm for the progressive pricing formulation gave near optimal solutions. In [24], Nikolaos *et al.* presented a mixed integer linear programming model of a home energy management system. A household including electric vehicles, DG such as roof top PV units, and energy storage systems along with controllable loads is considered. The system is proved to be efficient in minimizing the energy cost under dynamic pricing and reshaping the consumer load profile. Though the comfort associated with thermostatically controlled loads is included, inconvenience caused to the consumer due to scheduling of all controllable loads is not considered. Kim *et al.* [25] proposed the reinforcement learning (RL)-based algorithms for the dynamic pricing of the service provider and the energy consumption scheduling of a number of customers. The service provider and the customer exploit the learning capability in its decision making and an optimum schedule is obtained based on the observed retail price.

The inconvenience caused to the customer due to scheduling of the controllable loads is not modeled in this paper. Also, in all these research works, the authors do not consider the uncertain nature of the solar power in the models proposed. Algorithms for the residential DR considering user convenience under real-time and progressive pricing policies are presented by Laihyuk *et al.* [26]. The algorithm for the reformulated problem guarantee an optimal solution for the RTP policy.

From the literature, we can see that the future residential consumer will have renewable sources, schedulable loads, nonschedulable loads, and various time-dependent tariff. The power generated from renewable sources, RTP and arrival of nonschedulable loads are random. The consumer will want to minimize his electricity bill without compromising on the comfort. Thus, the problem is a decision making problem under uncertain environment, and RL [27] is an appropriate tool to solve such problems [28], [29]. The RL does not require explicit probabilistic models of the PV generation or controllable loads. It can learn from real data or a simulation model. We have used a simulation model in this paper. The RL can learn from real data such as actual PV generation in a home or actual loads with various degrees of constraints.

In this paper, the LCP model that can consider random renewable sources, random tariff, random critical load, and a smart way to capture the cost of delay in scheduling the loads is proposed. An RL-based approach to solve the LCP is also presented. It may be noted that, many of the models in the literature are special cases of the proposed model. The energy storage in terms of the battery will be an important part of future houses and will be considered in the future work.

The major contributions of this paper are listed as follows.

1) Development of a generalized mathematical model for the load scheduling problem including renewable sources.
2) Utilization of uncertainty modeling of PV generation using Beta probability density function (PDF) for load scheduling.
3) Development of the load scheduling algorithm using RL to minimize the electricity bill considering a renewable energy source.

This paper is organized as follows. In Section II, a generalized mathematical model is formulated. RL is explained in Section III. The RL algorithm for the load commitment without DG is discussed in Section IV. Modeling of the PV source with uncertainty is explained in Section V. In Section VI, load commitment with renewable DG is discussed. Simulation results, including the case study of the residential load commitment, are presented in Section VII. The conclusion is given in Section VIII.

II. Generalized Mathematical Model

Residential loads can be categorized into two groups, critical loads and controllable or flexible loads. Critical loads are must-run loads, which are being switched ON for a fixed period of time and the time of use cannot be shifted. Lighting loads, fan, TV, PC, etc., come under this category. For this type of loads, the customer should have the freedom to switch ON and OFF the loads as he desires. Controllable loads can be turned ON at any time slot within the given interval. Their operation can be delayed and/or interrupted if needed. They include cloth washers; cloth dryers; dish washers; heating, ventilating, and air-conditioning (HVAC) systems; water heaters; plug-in electric vehicles; battery chargers for consumer electronics; etc. Controllable loads can be further classified as atomic and nonatomic loads. Atomic loads are noninterruptible loads and once switched ON, will remain ON continuously for the specified duration. Nonatomic loads are interruptible loads and once turned ON can be turned OFF any number of times during the specified interval. Thus, the classification of loads can be generalized as critical, nonatomic, and atomic loads.

The load scheduling problem is formulated as an LCP. The objective of the LCP is to minimize the total cost of electricity consumed in a day subjected to various constraints. The LCP with and without renewable source is formulated.

For attaining a mathematical formulation, consider a consumer having m loads. A time line is defined such that one day consists of 24 equal time slots, k. The time between 12 AM and 1 AM is denoted as $k = 1$, the first time slot and $k = 24$ is the last time slot. Each individual load is modeled by a five-tuple represented as

$$d_j = (s_j, f_j, l_j, r_j, udc_j) \tag{1}$$

where $[s_j, f_j]$ represents the operating interval of load d_j. l_j denotes the total duration for which the load should remain ON. r_j is the power rating in kilowatt . If jth load is switched ON at s_j, it will remain ON from $k = s_j$ to $k = s_j + l_j - 1$.

Thus, $d_j = (5, 11, 3, 4)$ means a load d_j with 4-kW rating is to be switched ON, in between time slots 5 and 11, with a period of operation of three time slots. The problem is to obtain the optimum time slots at which the loads are to be switched ON, subject to different constraints.

udc_j is a parameter called the unit delay cost introduced to capture the degree of discomfort due to the delay in switching [22]. The switching ON time of a flexible load, e.g., a cloth washer can be delayed so as to operate it at a low-cost period. This delay causes some inconvenience or discomfort to the consumer. The parameter udc incorporated in the load model serves to capture the degree of discomfort. If the consumer is willing to accept delay, choose a low value of udc. Large value of udc indicates that the consumer cannot tolerate delay and results in a high energy cost. Later, in this paper, we will give guidelines for choosing udc.

A. With One Source of Energy, the Utility Grid

Let C^k denotes the energy cost per kilowatt hour for the kth time slot. It is assumed that the hourly tariff C^k is available day ahead in the case of ToUP or CPP. It is available one hour in advance in the case of the RTP.

Let u_j^k denotes the ON/OFF status of jth load in the kth time slot. If jth load is ON, $u_j^k = 1$ and if OFF $u_j^k = 0$, during kth time slot.

The LCP is to obtain a vector, $u = [u_1^1, u_1^2, \ldots, u_1^{24}, u_2^1, u_2^2, \ldots, u_2^{24}, \ldots, u_m^1, u_m^2, \ldots, u_m^{24}]$, so as to minimize the total electricity cost, $f(u)$ for a day.

Now, the LCP is as follows.

Minimize

$$\text{Total cost} = \sum_{k=1}^{24}\sum_{j=1}^{m} C^k r_j u_j^k \tag{2}$$

subject to

$$\sum_{k=s_j}^{f_j} u_j^k = l_j; \quad u_j^k = 0, \text{for} \quad k < s_j \quad \text{or} \quad k > f_j$$

$$\sum_{j=1}^{m} r_j u_j^k \leq \text{MDL}$$

where $k = 1 \ldots, 24, \quad j = 1, \ldots, m$, and MDL is the maximum demand limit for the kth slot.

The solution space is defined as

$$U = [u_1^1, u_1^2, \ldots, u_1^{24}, u_2^1, u_2^2, \ldots, u_2^{24}, \ldots, u_m^1, u_m^2, \ldots, u_m^{24}]$$

$$\forall u_j^k \in \{1, 0\}.$$

Then, the aim is to find the optimum values of u_j^k.

B. General Case With n Sources

Consider the case with n sources such as PV, grid, wind turbine, energy storage, etc. A general mathematical model for the load scheduling problem is to be developed to include all the sources.

Now, the load scheduling problem is to find a sequence of decisions $u = [u_{1i}^1, u_{1i}^2, \ldots, u_{1i}^{24}, u_{2i}^1, u_{2i}^2, \ldots, u_{2i}^{24}, \ldots, u_{mi}^1, \ldots, u_{mi}^{24}]$ $\forall i \in \{1, \ldots n\}$ $\forall u_{ji}^k \in \{0, 1\}$ such that the total electricity cost , $f(u)$ for a day is minimized.

The total cost function can be generalized as

$$\text{Total cost} = \sum_{k=1}^{24}\sum_{j=1}^{m}\sum_{i=1}^{n} C_i^k r_j u_{ji}^k \tag{3}$$

subject to

$$\sum_{i=1}^{n}\sum_{k=s_j}^{f_j} u_{ji}^k = l_j \ \forall j \in \{1, m\}$$

$$u_{ji}^k = 0, \text{for} \quad k < s_j \quad \text{or} \quad k > f_j$$

$$\sum_{j=1}^{m} r_j u_j^k \leq \text{MDL}$$

where C_i^k is the hourly energy cost of ith source, and u_{ji}^k is the binary status of jth load at kth hour, with ith source.

Here, $i = 1, 2, 3 \ldots n$ denotes different sources such as PV, grid, wind turbine, energy storage, etc.

III. RL

There exist a variety of machine learning strategies that takes the stochastic nature of the problem environment effectively. RL is a neurodynamic programming method that can give solutions to complex control problems. RL has been proposed for the solution of numerous control and optimization problems [30]. RL is also applied for solving the power system scheduling problems [31], [32]. In [22], an RL-based solution was proposed for DR. However, scheduling in the presence of DG is not considered. To make the paper self-contained, a brief overview of the RL algorithm for solving a general multistage decision making problem (MSDP) is first given [27].

Consider a system in state $x_0 \in \chi$ at stage 0 and an agent or decision maker take an action $a_0 \in A$ based on some policy π where, χ is the set of states and A is the set of permissible actions. The system state shifts to a new state x_1 depending on the action selected and property of the system. The system movement from x_0 to x_1, incurs a cost, $Co = g(x_0, a_0, x_1)$. In the RL literature, this cost is known as immediate reinforcement. In general, $g(x_0, a_0, x_1)$ can be a random variable. At any stage k, let the system be in state x_k, and the agent takes an action $a_k \in A$. Based on the action a_k and system characteristics, the system moves to a new state x_{k+1} and incurs a cost $g(x_k, a_k, x_{k+1})$. If the system possesses the Markov property, state transition at any stage k depends only on the system state at that stage, x_k and action taken a_k. Hence, $x_{k+1} = f(x_k, a_k)$. In general, $f(.,.)$ can be stochastic. In that case, the model of the system is completely specified by the transition probabilities $P_{x_k, x_{k+1}}^{a_k}$.

For the MSDP having T stages, the requirement is to find the optimal sequence of actions, $a_0, a_1, a_2, \ldots, a_{T-1}$, such that $E \sum_{k=1}^{N-1} \gamma g(x_k, a_k, x_{k+1})$ is minimized. Finding the sequence of optimal actions $a_0, a_1, a_2, \ldots, a_{T-1}$ is equivalent to finding a mapping from set of states to set of actions. This mapping π^* is called the optimal policy. (In the RL literature, any mapping $\pi : x- > a$ is called a policy.) Such a decision making problem can be solved using RL techniques even if transition probabilities are unknown. One of the RL algorithms is the Q-learning algorithm [27], [33], which is briefly explained next.

The Q-learning algorithm involves learning Q-values or more precisely optimal Q-values, $Q^*(x, a)$. If Q-value is mentioned without any qualifier, it is meant optimal Q-value. Optimal Q-values are represented as either $Q^*(x, a)$ or $Q(x, a)$. To understand the meaning of the optimal Q-value, let us first define Q-value under a policy π, denoted by $Q^\pi(x, a)$. It is defined as the expected total reinforcement if we start in state x, take an action a, and thereafter, follow policy π, i.e., $Q^\pi(x, a) = E \sum_{k=0}^{T-1} \gamma g(x_k, a_k, x_{k+1}) | x_0 = x, a_0 = a)$.

Optimal Q-values are the Q-values under the optimal policy π^*, i.e., $Q^*(x, a) = Q^{\pi^*}(x, a) = \min_\pi Q^\pi(x, a)$.

Therefore, optimal Q-value is the total expected cost if we start in state x, take an action a, and thereafter, follow an optimal policy. It has been shown in [27] that the optimal policy is given by $\pi^*(x) = \text{argmin}_{a \in \mathcal{A}} Q^*(x, a)$ and $\text{argmin}_{a \in \mathcal{A}} Q^*(x, a) = a^*$, if $Q(x, a^*) \leq Q(x, a) \, \forall a \in \mathcal{A}$. Thus, if Q-values are found for all the state action pairs, optimal policy can be retrieved. To find $Q^*(x, a)$, start with an initial guess $Q^0(x, a)$, for all state action pairs $Q(x, a)$. At each time instant k, the system is in state x_k, and we take an action a_k based on the current estimate of $Q^*(x_k, a_k)$ and $Q^n(x_k, a_k)$. If $Q^n(x_k, a_k)$ is a good approximation, we could have taken $a_g = \text{argmin}_{a \in \mathcal{A}} Q^n(x_k, a)$. This action is called the greedy action. In the initial part of the algorithm, $Q^n(x_k, a_k)$ will not be a good approximation of $Q^*(x_k, a_k)$. So, the issue is how to take actions while learning. One method is the ϵ-greedy algorithm for action selection. In this algorithm, the greedy action is chosen with probability $(1-\epsilon)$ and a random action with probability ϵ. It may be noted that if $\epsilon = 0$, the algorithm will always choose a greedy action, and if $\epsilon = 1$, the algorithm will always choose the random action. An optimum value for ϵ is chosen to get a good policy. Based on the action chosen, a new state x_k is reached and a cost of $g(x_k, a_k, x_{k+1})$ is incurred. Using this data, one can update the Q-value for the current state action pair using the following equation:

$$Q^{n+1}(x_k, a_k) = Q^n(x_k, a_k) + \eta[g(x_k, a_k, x_{k+1}) + \gamma \min_{a' \in \mathcal{A}_{x_k}} Q^k(x_{k+1}, a') - Q^k(x_k, a_k)]. \tag{4}$$

The update equation consists of two parameters, i.e., η and γ. Learning index η indicates how much the Q-values are modified at each of the learning steps. Discount factor γ indicates how much the future rewards are to be discounted. Updating of Q-value estimates are repeated a large number of times. If η is sufficiently small and if all possible (x, a) combinations of state and action are sufficiently visited, then the iteration given by (4) will result in Q^n converging to Q^* [27], [34].

IV. RL Algorithm for Load Commitment Without DG

In this section, a brief explanation of the RL algorithm for the LCP without DG is given [22]. As mentioned in Section III-A, the problem is to find a decision or commitment schedule $u_1, u_2, u_3, \ldots, u_m$, where u_j is a vector representing the status of the jth load. $u_j = [u_j^1, \ldots u_j^{24}]$.

The optimization problem is given by

$$\min_{u_j^k} \sum_{k=1}^{24} \sum_{j=1}^{m} C^k r_j u_j^k. \tag{5}$$

Since there are no coupling constraints, the aforementioned problem can be written as

$$\sum_{j=1}^{m} \left[\min_{u_j^k} \sum_{k=1}^{24} C^k r_j u_j^k \right]. \tag{6}$$

Now the problem is converted to m minimization problems, which can be solved independently.

Now, it is required to solve the following equation:

$$\min_{u_j^k} \sum_{k=1}^{24} C^k r_j u_j^k \tag{7}$$

subject to

$$\sum_{k=s_j}^{f_j} u_j^k = l_j \quad u_j^k = 0, \text{for} \quad k < s_j \quad \text{or} \quad k > f_j. \tag{8}$$

The aforementioned problem is modeled as a Markov decision process (MDP). To formulate the load scheduling problem as an MDP and to make use of RL [27], state, state space, transition function, action, and reward function are to be identified with respect to the scheduling problem. The information needed to arrive at the decision should be contained in the state of the system. So, in the case of load scheduling problem, the state at any stage k is represented by the vector $[x(1), x(2)]$, where $x(1)$ is the current time slot and $x(2)$ is the total ON duration of the load.

The MDP starts with $x(1) = k = s_j$ and $x(2) = 0$ and terminates when $x(1) = f_j$ or $x(2) = l_j - 1$.

The state space is given by

$$\chi = \{(x(1), x(2));\ x(1) \in s_j, \ldots, f_j, x(2) \in 0, 1, \ldots, l_j\}.$$

At each state, the action to be taken is whether to switch ON or OFF the load. The action set $\mathcal{A} = \{0, 1\}$.

The transition from the current state to the next state by the application of an action is defined by the transition function.

From the present state, the new state is obtained using the equation

$$[x_{k+1}(1),\ x_{k+1}(2)] = [x_k(1) + 1,\ x_k(2) + a]. \tag{9}$$

The objective is to minimize the cost of electricity subject to the constraints. To capture this objective, the cost incurred when the process moves from x to x_{new} can be used to formulate the reward function, $g(x, a, x_{\text{new}})$.

Therefore, the reward function is given by

$$\begin{aligned} g(x, a, x_{\text{new}}) &= C^k r_j; \quad \text{if } a = 1 \\ &= udc_j r_j; \quad \text{if } a = 0 \\ &= \text{penality}; \quad \text{if} \\ x_{\text{new}}(1) &= f \text{ and } x_{\text{new}}(2) < l. \end{aligned} \tag{10}$$

In order to incorporate the PV power with load scheduling, first a suitable stochastic model is to be developed for PV. The stochastic modeling of PV is discussed in the next section.

V. Uncertainty Modeling of the PV System

The output of the renewable DG unit like PV source is stochastic due to the uncertain nature of the solar irradiance. For simulation studies, the hourly solar irradiance is to be modeled taking into account the uncertainties associated with it.

So, a stochastic model is needed for considering the random behavior of these sources. A lot of previous data of irradiance is available, and for developing the descriptive model for the same, Beta PDF is being used [35].

$$\begin{aligned} f_b(s) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} * (s)^{\alpha-1} * (1-s)^{\beta-1} \\ &\qquad \text{for } 0 \le s \le 1, \alpha, \beta \ge 0 \\ &= 0, \qquad \text{otherwise} \end{aligned} \tag{11}$$

where

s solar irradiance kW/m^2;
$f_b(s)$ Beta distribution function of s;
α, β parameters of Beta distribution function.

α and β are calculated from the mean (μ) and standard deviation (σ) of the random variable s as

$$\beta = (1-\mu) * \left(\frac{\mu * (1+\mu)}{\sigma^2} - 1\right) \tag{12}$$

$$\alpha = \frac{\mu * \beta}{1-\mu}. \tag{13}$$

To model the hourly solar irradiance using Beta PDF, we need the mean and standard deviation of the historical data. This is obtained by historical data processing. The hourly solar irradiance of the site under study for a period of one year is considered. The year is divided into four seasons and each season with 3 months, is represented by any day in that season. The day is further divided into 24-h segments. There will be 90 irradiance for each hour in a season, considering 30 days for a month (3 months × 30 days). For each hour, the mean and standard deviation of the solar irradiance is computed from this data and Beta pdf is generated [35].

The power generation of the PV module depends on the solar irradiance, ambient temperature, and the module characteristics. For a particular hour, solar irradiation "*s*" is generated using the corresponding Beta PDF.

Then, the PV output power P_S, is calculated using the following equations:

$$\text{FF} = \frac{V_{\text{MPP}} * I_{\text{MPP}}}{V_{\text{oc}} * I_{\text{sc}}} \tag{14}$$

$$T_c = T_A + s\left(\frac{N_{\text{OT}} - 20}{0.8}\right) \tag{15}$$

$$I_y = s[I_{\text{sc}} + K_i(Tc - 25)] \tag{16}$$

$$V_y = V_{oc} - K_v * T_c \tag{17}$$

$$P_S = N * FF * V_y * I_y. \tag{18}$$

To obtain the Beta PDF parameters, the hourly irradiance data available for a particular site can be used. The site can be located by specifying the latitude and longitude. The computed parameters are used for generating the Beta PDF for each hour and the hourly solar irradiance can be simulated. The PV module parameters are also known. The power output of the PV module can be obtained using (18). The PV power output is utilized for load scheduling with DG.

VI. RL Algorithm for Load Commitment With Renewable DG

In this section, algorithm for the LCP with DG is developed. It is assumed that the grid-connected home is equipped with a roof-top PV system. The task is to minimize the daily electricity cost considering the two sources, the PV and the grid, subject to the constraints.

That is

$$\min_{u_{ji}^k} E\left[\sum_{k=1}^{24}\sum_{i=1}^{2} C_i^k r_j u_{ji}^k\right] \tag{19}$$

subject to

$$\sum_{k=s_j}^{f_j} u_{ji}^k = l_j$$

$$u_{ji}^k = 0, \text{for} \quad k < s_j \quad \text{or} \quad k > f_j. \tag{20}$$

Here, $i = 1$ for the grid and $i = 2$ for PV. u_{ji}^k indicates the status of load j, in the time slot k. That is, $u_{ji}^k = 1/0$ denotes the ON/OFF status. E[] denotes the expected value.

The state is defined same as, $[x(1), x(2)]$, where $x(1)$ is the current time slot and $x(2)$ is the total ON duration of the load. At each state, the action to be taken is whether to switch ON, from the grid/PV or switch OFF the load. The action set is modified to $\mathcal{A} = \{0, 1, 2\}$. For switching OFF, the action is denoted as, $a = 0$. Action a is taken as 1 for switching ON the device using the grid power. When the load is switched ON from the PV power, $a = 2$. This action causes a reduction in the available PV power P_{pv}, during that particular hour and the PV power is updated to, P_{pvnew} given by

$$P_{\text{pvnew}} = P_{\text{pv}} - r_j. \tag{21}$$

In the RL approach, the stochastic nature of the PV power can be incorporated in the reward function and the reinforcement function is defined as

$$\begin{aligned} g(x, a, x_{\text{new}}) &= C_i^k r_j; \quad \text{if } a = 1 \text{ or } 2 \\ &= \text{penality}; \ \text{if } P_{\text{pvnew}} < r_j \text{ and } a = 2 \\ &= udc_j r_j; \quad \text{if } a = 0 \\ &= \text{penality}; \\ &\quad \text{if } x_{\text{new}}(1) = f \text{ and } x_{\text{new}}(2) < l \end{aligned} \tag{22}$$

where C_i^k denotes the cost of one unit of energy during the kth time slot for ith source. C_i^k depends on the nature of power source i, solar or grid and also on the type of tariff in the case of the grid power as mentioned in Section II. It is to be noted that, after load scheduling, the PV power availability at that time slot is updated to P_{pvnew}. If the power available is less than r_j, a penalty is assigned as reward function. This is the mechanism by which the algorithm communicates the desired goal. State transition function is defined as

$$[x_{k+1}(1),\ x_{k+1}(2)] = [x_k(1) + 1,\ x_k(2) + a'] \tag{23}$$

where $a' = 0$, if the device is in OFF state, i.e., $a = 0$ and $a' = 1$, if the device is switched ON either from the grid or PV, i.e., $a = 1$ or $a = 2$.

The Q-Learning using RL involves two phases, learning and running. In the learning phase, Q-values of different state-action pairs are updated. The state in this case contains the information regarding the index of load, time slot, and ON duration. Before starting Q-value updation, the α and β parameters are generated from the historical data. Also the $(s_j, f_j, l_j, r_j, udc_j)$ data for each load is made available. The Q-values are initialized to zero for each load, i.e., $Q^0(x, a) = 0$, where state $x = [x(1)\ x(2)]$, $a = \{0, 1, 2\}$. The PV power for the time slot is simulated using Beta PDF, which takes into account the randomness of the PV

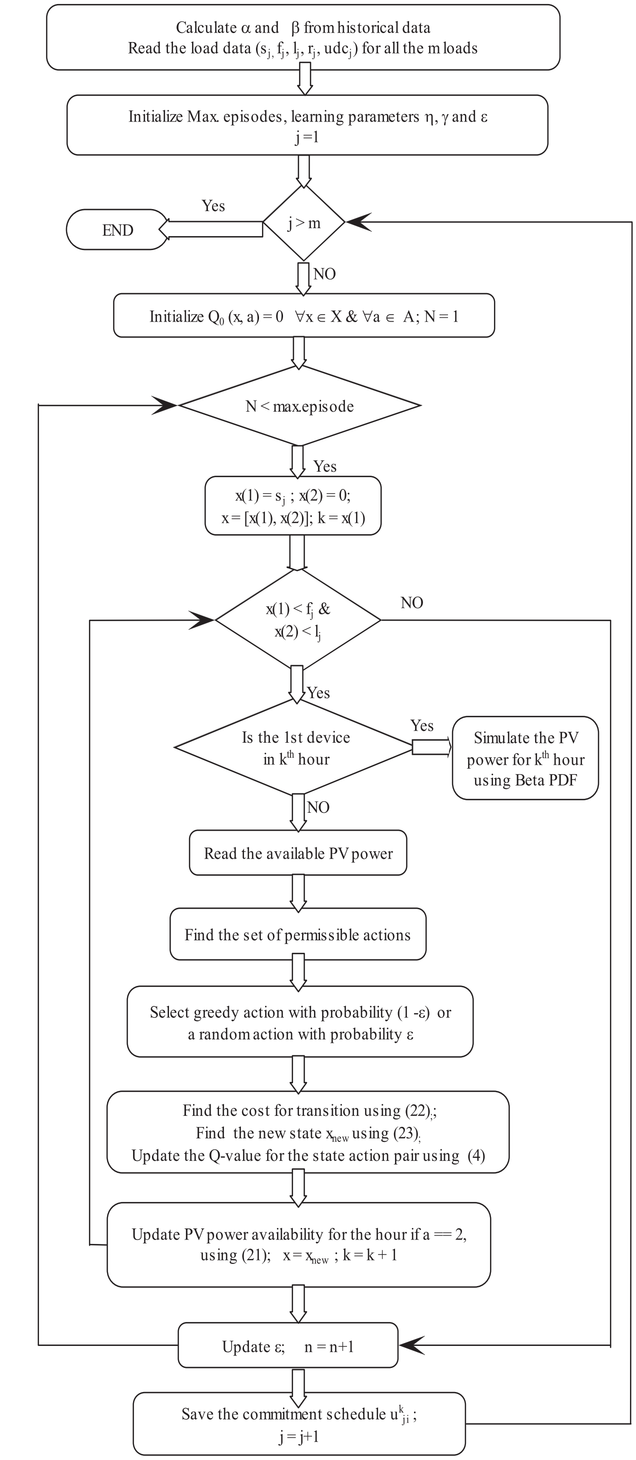

Calculate α and β from historical data
Read the load data (s_j, f_j, l_j, r_j, udc_j) for all the m loads
Initialize Max. episodes, learning parameters η, γ and ε
j = 1
j > m
Yes
END
NO
Initialize Q_0 (x, a) = 0 ∀x ∈ X & ∀a ∈ A; N = 1
N < max.episode
Yes
x(1) = s_j ; x(2) = 0;
x = [x(1), x(2)]; k = x(1)
x(1) < f_j &
x(2) < l_j
NO
Yes
Is the 1st device
in k^th hour
Yes
Simulate the PV
power for k^th hour
using Beta PDF
NO
Read the available PV power
Find the set of permissible actions
Select greedy action with probability (1 -ε) or
a random action with probability ε
Find the cost for transition using (22).;
Find the new state x_new using (23).
Update the Q-value for the state action pair using (4)
Update PV power availability for the hour if a == 2,
using (21); x = x_new ; k = k + 1
Update ε; n = n+1
Save the commitment schedule u^k_ji ;
j = j+1

Fig. 1. Flowchart for the RL algorithm.

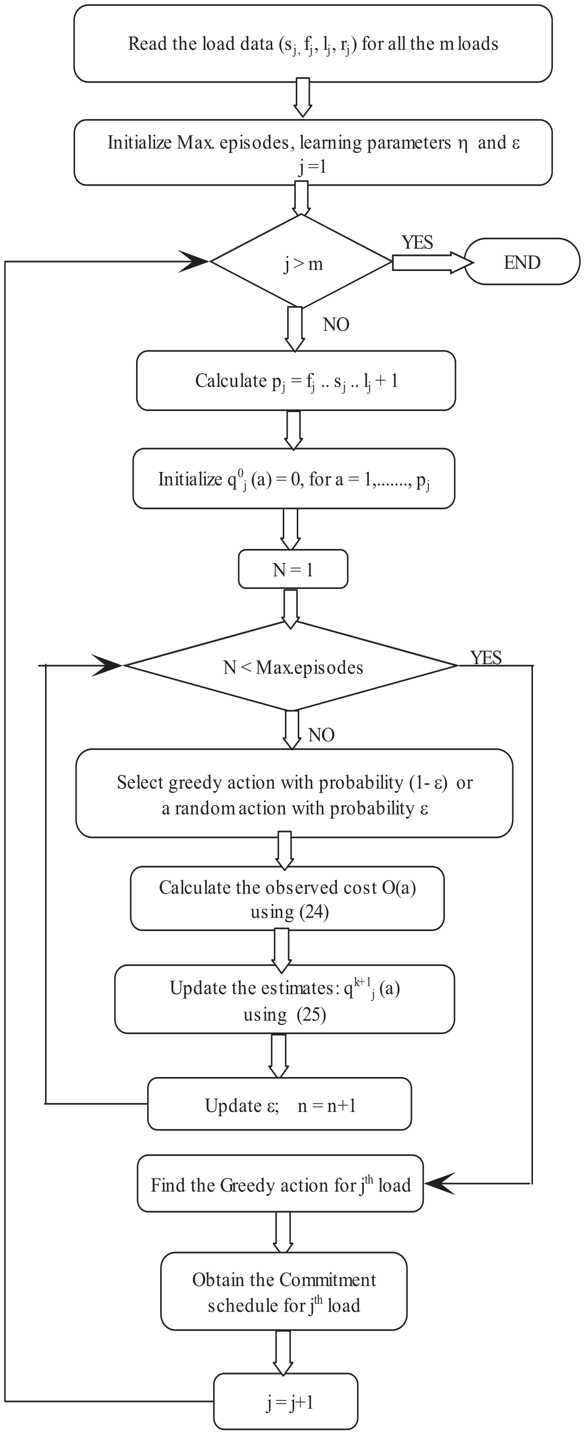

Read the load data (s_j, f_j, l_j, r_j) for all the m loads
Initialize Max. episodes, learning parameters η and ε
j = 1
j > m
YES
END
NO
Calculate p_j = f_j .. s_j .. l_j + 1
Initialize q^0_j (a) = 0, for a = 1,......., p_j
N = 1
N < Max.episodes
YES
NO
Select greedy action with probability (1- ε) or
a random action with probability ε
Calculate the observed cost O(a)
using (24)
Update the estimates: q^k+1_j (a)
using (25)
Update ε; n = n+1
Find the Greedy action for j^th load
Obtain the Commitment
schedule for j^th load
j = j+1

Fig. 2. Flowchart for the LA algorithm.

power. Starting from $x(1) = s_j$ and $x(2) = 0$, action/decision is taken based on the availability of the PV power following the ϵ-greedy strategy. Depending on the action taken, state transition occurs and the Q-value as well as the PV power availability are updated. After sufficient number of iterations, the learning phase converges, which is decided by the negligible updation in Q-values. The simulation phase gives the best action or source switching of each load corresponding to different time slots. The steps involved in the RL algorithm are shown in Fig. 1 as a flowchart.

We compare our algorithm with the learning automata (LA) algorithm, which is used for solving a single-stage decision-making problem (SSDP). If we consider only atomic loads, the LCP will become an SSDP. So to make the paper self-contained, we give a brief explanation about the LA algorithm here [36].

The load scheduling problem with m loads is viewed as m independent single stage decision-making problems. The obser-

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
PV MODULE PARAMETERS

Watt peak(W)	75
Open circuit voltage(V)	21.98
Short circuit current (A)	5.32
Voltage at maximum power(V)	17.32
Current at maximum power(V)	4.76
Voltage temperature coefficient(mV/^{0}C)	14.4
Current temperature coefficient(mA/^{0}C)	1.22
Nominal cell operating temperature(^{0}C)	43

TABLE II
LOAD DETAILS

Load	s	f	l	r	*udc*
1	9	17	3	5	1
2	8	15	4	3	1
3	11	19	5	7	2
4	10	24	7	5	1
5	20	24	2	5	3
6	6	18	8	5	1

vation or reward on taking an action "*a*" is given by

$$O(a) = \sum_{k=s_j+a}^{s_j+a+l_j-1} C^k r_j. \tag{24}$$

The objective is to find the best action a^* that minimizes the expected value of the reward $q(a) = E\{O((a)\}$, which can be obtained using the following equation:

$$q_j^{k+1}(a) = q_j^k(a) + \eta(O(a) - q_j^k(a)). \tag{25}$$

The flowchart for the LA algorithm is shown in Fig. 2.

VII. SIMULATION RESULTS

A. Modeling the Uncertainty of the PV Power

The output power of the PV module is dependent on the solar irradiance, the characteristics of the module and ambient temperature. The uncertainty in the solar irradiance is modeled using Beta PDF as stated earlier. The parameters of the Beta PDF of solar irradiation is estimated from the hourly irradiance data taken from the National Renewable Energy Lab, India, for Thrissur, Kerala (Latitude 10.52°N and Longitude 76.21°E). The four seasons in Kerala can be divided as winter, summer, south west monsoon, and retreating monsoon. There are 90 data points corresponding to a season. The Beta PDF for each hour is generated and using this the hourly irradiance is simulated. The PV source is designed to meet about 50% of the total load for 24 h. The characteristics of the PV module [35] is given in Table I.

B. Load Scheduling

To test the efficacy and scalability of the developed algorithm for load scheduling with DG using RL, several simulation experiments were conducted. Max_episodes and the learning parameters η and ϵ are initialized as 1000, 0.1, and 0.5, respectively. ϵ is updated to $0.9 * \epsilon$ after every 100 episodes. Initially, a system consisting of six loads is considered to test the efficacy of the algorithm and the load characteristics is given in Table II. The price of the PV power is assumed to be a small value compared to the price of the grid power. The price of the electricity from the grid is given in Table III. The simulation platform employed is MATLAB and a computer with the following configuration is used for obtaining the results: Intel core i5, clocking at 3.47 GHz and 4-GB RAM.

TABLE III
PRICE OF ELECTRICITY

Time,k	Price,C_k
1...12	5
13,14	12
15...18	5
19,20,21	10
22,23,24	5

TABLE IV
SCHEDULE WITHOUT DG

Hour	Load 1	Load 2	Load 3	Load 4	Load 5	Load 6	Demand from grid
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0
6	0	0	0	0	0	1	5
7	0	0	0	0	0	1	5
8	0	1	0	0	0	1	8
9	1	1	0	0	0	1	13
10	1	1	0	1	0	0	13
11	1	1	1	1	0	1	25
12	0	0	1	1	0	1	17
13	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0
15	0	0	1	1	0	1	17
16	0	0	1	1	0	1	17
17	0	0	1	1	0	0	12
18	0	0	0	1	0	0	5
19	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0
22	0	0	0	0	1	0	5
23	0	0	0	0	1	0	5
24	0	0	0	0	0	0	0

TABLE V
EFFECT OF udc ON THE COST

udc (for all loads)	cost
0	735
3	735
5	795
7	890
8	988
20	988

1) Load Scheduling Without DG: The simulation is done without PV source first. All the loads are scheduled in the low-priced period within their specified time slots with a minimum total energy cost of 735 units as shown in Table IV. With same set of loads, simulations were repeated with different values of udc ranging from 0 to 20, to study the effect of udc. It is observed that the total cost of energy increased with udc. The results are tabulated in Table V. From the table, we can see that when udc

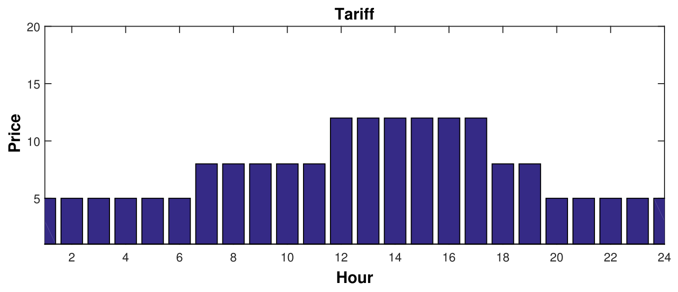


Fig. 3. Tariff.

TABLE VI
PERFORMANCE COMPARISON WITHOUT PV

Algorithm	Scheduled Cost	Unscheduled Cost	Computation Time(s)
RL	5014	5161	3.89
LA	5004	5055	6.76

increases, the cost increases as consumer convenience is given importance. Moreover, we can also see that when udc is above 7, there is little impact. Here, the daily average tariff of electricity is around seven units. From this, we can give guidelines for choosing udc. If convenience is of utmost importance choose udc greater than the average tariff (C_{av}) and choose a value between 0 and C_{av} if delay can be tolerated.

To demonstrate the scalability and validate the same, 100 random loads are generated, assuming $udc = 0$ for all loads. Five categories of loads based on the time of occurrence are considered. Category 1 is assumed to be nonflexible loads that can occur at any time of the day with a duration of 1–4 h. The other categories include flexible loads occurring at the load peak of the day, load peak at night, and off-peak hours. So, category II load is with s_j randomly selected between the 8th and 15th interval and the scheduling window is randomly selected between one and seven intervals. The length is chosen anywhere between one and three intervals. Category III load is chosen such that it starts between 17th and 20th interval. The scheduling window is randomly chosen between one and five intervals and the length is between one and two intervals. Category IV load is assumed to occur between 1st and 8th interval, the off-peak hours after midnight. The length is chosen anywhere between one and three intervals and the scheduling window is randomly chosen between one and five intervals. Category V is with a start time randomly selected between the 13th and 15th interval, a scheduling window of one to five intervals and length between one and three intervals. A load generator is programmed to generate the loads. The algorithm is validated and compared with the LA algorithm proposed in [36]. The differential tariff used is shown in Fig. 3. In the absence of PV, all loads are committed in the low-priced time slots and the total price of the electricity for the learned schedule is 5 014 units. With the unscheduled load, the total price of the electricity is found to be 5 161 units. The computation time of RL without PV for 100 loads is found to be 3.89 s, which is less than that of the LA algorithm. The performance comparison of the algorithms without PV is given in Table VI. It can be seen that there is small variation in the costs, from that of the LA algorithm and this is due to the random generation of the 100 loads. The LA is tested with a set of random generated loads with an unscheduled cost of 5 055. When we generate the load a second time using the program, there may be slight variation in the load profile and resulted in an unscheduled cost of 5 161.

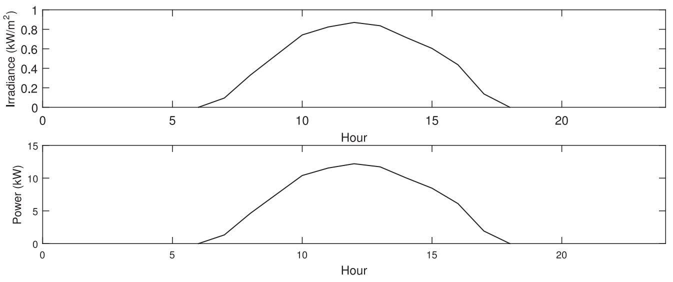


Fig. 4. Simulated irradiance and power.

TABLE VII
SCHEDULE WITH DG

Hour	Load 1	Load 2	Load 3	Load 4	Load 5	Load 6	Demand from grid
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0
6	0	0	0	0	0	1	5
7	0	0	0	0	0	1	5
8	0	2	0	0	0	1	5
9	2	0	0	0	0	1	5
10	2	2	0	1	0	1	10
11	2	2	1	1	0	1	17
12	0	2	2	1	0	1	10
13	0	0	2	0	0	0	0
14	0	0	2	0	0	0	0
15	0	0	2	1	0	1	10
16	0	0	0	2	0	0	0
17	0	0	0	1	0	0	5
18	0	0	0	1	0	0	5
19	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0
22	0	0	0	0	1	0	5
23	0	0	0	0	1	0	5
24	0	0	0	0	0	0	0

2) Load Scheduling With Renewable DG: Next, simulation is done with the PV source. As mentioned earlier, to simulate the random solar irradiation, we use the Beta distribution function. Hence, while learning in each iteration, the algorithm "sees" different PV power, and in the simulations, tested for each day the PV power will be different. One such scenario is shown in Fig. 4. The commitment schedule obtained using a simulated power scenario is given in Table VII. In the schedule, the ON status of the load with the power taken from the PV source is denoted as 2 and that with the grid power is denoted as 1. All the loads are scheduled between their respective s_j's and f_j's in the low-priced period. At any time slot, when the PV power is sufficient to meet the load demand, the scheduling is done using the PV power. The remaining loads are committed with the grid power in the low-priced period and satisfying the constraints. For example, load 1 is committed during the interval $\{9–11\}$ and is powered by the PV source. Load 3 is ON during $\{11–15\}$, with

TABLE VIII
PARAMETERS OF HOUSEHOLD DEVICES

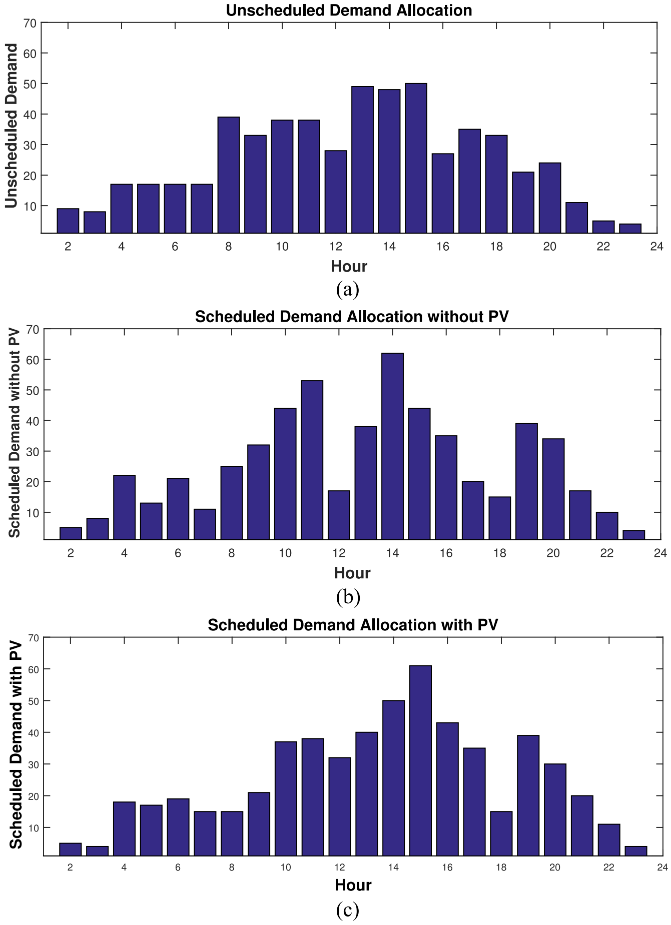
Sl. No.	Appliance	s	f	l	r	*udc*
1	Lighting	18	20	3	0.36	8
2	Washing machine	8	14	2	0.50	1
3	Cloth dryer	11	17	1	1.80	0
4	Dish washer	14	19	2	1.20	9
5	Vacuum cleaner	9	14	1	0.65	1
6	Iron	6	9	1	1.10	1
7	Rice cooker	7	10	1	0.30	0

grid power for the 11th slot and PV power for slots {12–15}. It can be seen that slot 11 is a low-priced period for the grid power. For load 4, the schedule is {10–12} and {15–18} with the 16th slot powered from PV. With DG, the total energy cost is reduced to 495 and the demand from the grid is also reduced. The simulations are repeated for a large number of scenarios. Even though there is variation in the simulated PV power due to uncertainty, it is observed that the load schedules obtained corresponds to the minimum energy cost.

Next, to demonstrate the scalability and validate the same, 100 random loads are generated, assuming $udc = 0$ for all loads. The PV source capacity is enhanced to account for the increased number of loads. The simulation is carried out with the same loads. In this case, the price of electricity is reduced to 2 130 units. The unscheduled load is shown in Fig. 3(a). The load scheduled using the RL algorithm, without PV and with PV are shown in Fig. 3(b) and (c), respectively. It can be seen that without PV, as the load tries to avoid periods with higher price, though there is reduction in energy cost, the maximum demand on the system is increased. The results show that, with PV, there is reduction in the energy cost and grid power consumption. The computation time in the case of RL with PV is increased to 563 s due to the increase in number of actions, which results in more number of Q-values to be stored.

C. Case Study

To investigate the performance of the proposed scheduling algorithm, a home with schedulable and unschedulabe appliances and PV source is considered. Typical schedulable appliances in the home include washing machine, cloth dryer, dish washer, vacuum cleaner, iron, and rice cooker. These appliances operate with different time intervals, duration of operation, and power ratings. Also, the consumer's willingness to tolerate delay in operation of the schedulable devices is reflected through the choice of udc value. The parameters of the household devices are given in Table VIII. The dish washer assigned with $udc = 9$ indicates that the consumer is not willing to accept the inconvenience caused by delay. We cannot shift the time of operation of a nonflexible load such as lighting load. The duration of operation is 3 h starting from 18th time slot. The study is first conducted without considering the DG source. As expected, the appliances are committed satisfying all the specified requirements with an energy cost of 53.65. The appliance schedule is given in Table IX. If the consumer is willing to tolerate delay for the operation of the dish washer, then a low value of udc can be set for this load. Simulation is repeated with $udc = 0$ for the dish washer load. It was observed that the appliance schedule shifted to the low-priced periods with a minimum total energy cost of 45.25.



(a)

(b)

(c)

Fig. 5. (a) Unscheduled demand. (b) Scheduled demand without PV using RL. (c) Scheduled demand with PV using RL.

For the same residential loads given in Table VIII, simulation is also done with the binary particle swarm optimization (BPSO) algorithm proposed in [37]. The results are compared with the results obtained using the RL algorithm and is shown in Table X. It is observed that the scheduled cost obtained in both the cases is the same, minimum cost of 45.45. But the computation time is less in the case of the RL algorithm (1.34 s) compared to the BPSO algorithm (16.51 s). This is because the computation time in the BPSO algorithm depends on the binary particle size and takes more time to converge. With PV source and comfort constraint, the LCP is more complex and it is difficult to obtain solution with BPSO.

The PV generation of the home considered is 5 kW. The appliance commitment schedule with DG is given in Table XI. In the schedule, the ON status of load with the power taken from the PV source is denoted as 2 and that with the grid power is denoted as 1. It can be seen that in this case for the lighting load, which is unschedulable and required to be operated in time slots

TABLE IX
HOME APPLIANCE SCHEDULE WITHOUT DG

Hour	Lighting	Washing machine	Cloth dryer	Dish washer	Vacuum cleaner	Iron	Rice cooker	Grid Power
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	1	0	1.1
7	0	0	0	0	0	0	1	0.3
8	0	1	0	0	0	0	0	0.5
9	0	1	0	0	1	0	0	1.15
10	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0
12	0	0	1	0	0	0	0	1.8
13	0	0	0	0	0	0	0	0
14	0	0	0	1	0	0	0	1.2
15	0	0	0	1	0	0	0	1.2
16	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0
18	1	0	0	0	0	0	0	0.36
19	1	0	0	0	0	0	0	0.36
20	1	0	0	0	0	0	0	0.36
21	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0

TABLE X
PERFORMANCE COMPARISON WITH BPSO

Scheduled Cost	Unscheduled Cost	Computation Time(s) with RL	Computation Time (s) with PSO
45.45	53.65	1.34	16.51

TABLE XI
HOME APPLIANCE SCHEDULE WITH DG

Hour	Lighting	Washing machine	Cloth dryer	Dish washer	Vacuum cleaner	Iron	Rice cooker	Grid Power
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	2	0
8	0	2	0	0	0	2	0	0
9	0	2	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0
13	0	0	0	0	2	0	0	0
14	0	0	0	2	0	0	0	0
15	0	0	0	2	0	0	0	0
16	0	0	2	0	0	0	0	0
17	0	0	0	0	0	0	0	0
18	1	0	0	0	0	0	0	0.36
19	1	0	0	0	0	0	0	0.36
20	1	0	0	0	0	0	0	0.36
21	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0

$\{18–20\}$, power is taken from the grid, due to the nonavailability of the PV power at that time slots. Also, the dish washer load is assigned a high udc to avoid delay in scheduling. Consequently, it is ON during time slots $\{14–15\}$. With DG, the energy cost is reduced to 16.25.

VIII. CONCLUSION

We have presented a unified and implementable formulation for the load scheduling problem like that of the UCP and clearly spelled out the assumptions made. The discomfort caused due to delay in scheduling a load is also considered in the load model by including a factor udc. We have given precise guidelines to choose this parameter. The PV type of the renewable DG is becoming popular because of its abundance. The output power of the PV source is stochastic in nature due to the uncertainty associated with the solar irradiance. In this paper, the random nature of the solar irradiance is modeled using Beta PDF. We have also presented an algorithm based on the RL approach for the LCP. A large number of simulations are done and the results obtained are analyzed and verified. From the simulation results, we found that the loads are scheduled with the minimum energy cost, effectively utilizing the PV and grid, and satisfying the constraints that will benefit the consumer. The formulation can be made more general by including the storage within the consumer premises into the model.

REFERENCES

[1] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation and Control*. New York, NY, USA: Wiley, 2003.

[2] C. W. Gellings and W. M. Smith, "Integrating demand-side management into utility planning," *Proc. IEEE*, vol. 77, no. 6, pp. 908–918, Jun. 1989.

[3] B. R. Parekh, A. T. Davda, B. Azzopardi, and M. D. Desai, "Dispersed generation enable loss reduction and voltage profile improvement in distribution network-case study, Gujarat, India," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1242–1249, May 2014.

[4] J. M. Carrasco *et al.*, "Power-electronic systems for the grid integration of renewable energy sources: A survey," *IEEE Trans. Ind. Electron.*, vol. 53, no. 4, pp. 1002–1016, Jun. 2006.

[5] S. Kouro, J. I. Leon, D. Vinnikov, and L. G. Franquelo, "Grid-connected photovoltaic systems: An overview of recent research and emerging PV converter technology," *IEEE Ind. Electron. Mag.*, vol. 9, no. 1, pp. 47–61, Mar. 2015.

[6] M. H. Albadi and E. F. El-Saadany, "A summary of demand response in electricity markets," *Electr. Power Syst. Res.*, vol. 78, no. 11, pp. 1989–1996, 2008.

[7] I. Ahamed, T. P. Syed, Q. Ali, and N. H. Malik, "Learning automata algorithms for load scheduling," *Electr. Power Compon. Syst.*, vol. 41, pp. 286–303, 2013.

[8] A.-H. Mohsenian-Rad, V. W. S. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid," *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 320–331, Dec. 2010.

[9] J. Aghaei and M.-I. Alizadeh, "Demand response in smart electricity grids equipped with renewable energy sources: A review," *Renewable Sustain. Energy Rev.*, vol. 18, pp. 64–72, 2013.

[10] M. Muratori and G. Rizzoni, "Residential demand response: Dynamic energy management and time-varying electricity pricing," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1108–1117, Mar. 2016.

[11] M. H. J. Weck, J. Hooff, and W. G. J. H. M. Sark, "Review of barriers to the introduction of residential demand response: A case study in the Netherlands," *Int. J. Energy Res.*, vol. 41, no. 6, pp. 790–816, 2017.

[12] D. S. T. Logenthiran and T. Z. Shun, "Demand side management in smart grid using heuristic optimization," *IEEE Trans. Smart Grid*, vol. 3, pp. 1244–1252, Sep. 2012.

[13] E. Gilboa, P. Yang, P. Chavali, and A. Nehorai, "Parallel load schedule optimization with renewable distributed generators in smart grids," *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1431–1441, Sep. 2013.

[14] S. Zhou, Z. Wu, J. Li, and X.-p. Zhang, "Real-time energy control approach for smart home energy management system," *Electr. Power Compon. Syst.*, vol. 42, no. 3–4, pp. 315–326, 2014.

[15] C. O. Adika and L. Wang, "Autonomous appliance scheduling for household energy management," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 673–682, Mar. 2014.

[16] Y. Guo, M. Pan, and Y. Fang, “Optimal power management of residential customers in the smart grid,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 9, pp. 1593–1606, Sep. 2012.

[17] X. F. B. Ruan, Q. Yang, and W. Yan, “Demand response under real-time pricing for domestic energy system with DGs,” in *Proc. Int. Conf. Power Syst. Technol.*, 2014.

[18] D. Setlhaolo, X. Xia, and J. Zhang, “Optimal scheduling of household appliances for demand response,” *Electr. Power Syst. Res.*, vol. 116, pp. 24–28, 2014.

[19] T. Li and M. Dong, “Real-time residential-side joint energy storage management and load scheduling with renewable integration,” *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 283–298, Jan. 2018.

[20] S. L. Arun and M. P. Selvan, “Intelligent residential energy management system for dynamic demand response in smart buildings,” *IEEE Syst. J.*, vol. 12, no. 2, pp. 1329–1340, Jun. 2018.

[21] A. Anvari-Moghaddam, H. Monsef, and A. Rahimi-Kian, “Optimal smart home energy management considering energy saving and a comfortable lifestyle,” *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 324–332, Jan. 2015.

[22] T. P. I. Ahamed, S. D. Maqbool, and N. H. Malik, “A reinforcement learning approach to demand response,” in *Proc. Centenary Conf. Electr. Eng.*, Indian Institute of Science, Bangalore, India, 2011, pp. 168–172.

[23] M. Castillo-Cagigal and E. Matallanas, “Neural network controller for active demand side management with PV energy in the residential sector,” *Appl. Energy*, vol. 91, pp. 90–97, 2012.

[24] N. G. Paterakis, O. Erdinc, A. G. Bakirtzis, and J. P. S. Catalão, “Optimal household appliances scheduling under day-ahead pricing and load-shaping demand response strategies,” *IEEE Trans. Ind. Inf.*, vol. 11, no. 6, pp. 1509–1519, Dec. 2015.

[25] B.-G. Kim, Y. Zhang, M. van der Schaar, and J.-W. Lee, “Dynamic pricing and energy consumption scheduling with reinforcement learning,” *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2187–2198, Sep. 2016.

[26] L. Park, Y. Jang, S. Cho, and J. Kim, “Residential demand response for renewable energy resources in smart grid systems,” *IEEE Trans. Ind. Inf.*, vol. 13, no. 6, pp. 3165–3173, Dec. 2017.

[27] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998.

[28] R. H. Crites and A. G. Barto, “Elevator group control using multiple reinforcement learning agents,” *Mach. Learn.*, vol. 33, no. 2–3, pp. 235–262, 1998.

[29] F. Sahba, H. R. Tizhoosh, and M. M. A. Salama, “Application of reinforcement learning for segmentation of transrectal ultrasound images,” *BMC Med. Imag.*, vol. 8, no. 1, p. 1–10, 2008.

[30] G. Tesauro, “Temporal difference learning and TD-Gammon,” *Commun. ACM*, vol. 38, no. 3, pp. 58–68, 1995.

[31] D. Ernst, M. Glavic, and L. Wehenkel, “Power systems stability control: Reinforcement learning framework,” *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 427–435, Feb. 2004.

[32] L. Xiao, X. Xiao, C. Dai, M. Pengy, L. Wang, and H. V. Poor, “Reinforcement learning-based energy trading for microgrids,” arXiv:1801.06285, 2018.

[33] D. P. Bertsekas and J. N. Tsitsiklis, “Neuro-dynamic programming: An overview,” in *Proc. IEEE 34th Conf. Decis. Control*, vol. 1, 1995, pp. 560–564.

[34] M. A. L. Thathachar and P. S. Sastry, *Networks of Learning Automata: Techniques for Online Stochastic Optimization*. New York, NY, USA: Springer, 2003.

[35] M. M. A. Salama Y. M. Atwa, E. F. EI-Saadany, and R. Seethapathy, “Optimal renewable resources mix for distribution system energy loss minimization,” *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 360–370, Feb. 2010.

[36] T. P. Imthias Ahamed, S. Q. Ali, and N. H. Malik, “Learning automata algorithms for load scheduling,” *Electr. Power Compon. Syst.*, vol. 41, no. 3, pp. 286–303, 2013.

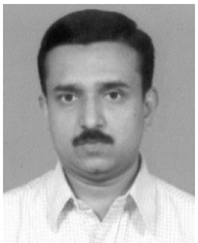
[37] Remani T., E. A. Jasmin, and T. P. I. Ahamed, “Load scheduling with maximum demand using binary particle swarm optimization,” in *Proc. Int. Conf. Adv. Power Energy*, 2015, pp. 294–298.

Remani T. received the B. Tech. degree in electrical engineering from Kerala University, Thiruvananthapuram, India, in 1985, and the M. Tech. degree in instrumentation and control systems from the National Institute of Technology Calicut, Kozhikode, India, in 1988. She is currently working toward the Ph. D. degree with the Department of Electrical Engineering, Government Engineering College, Thrissur, India.

She has 28 years of teaching experience. Her current research interests include power system operation and control, demand-side management, demand response, etc.

E. A. Jasmin received the bachelor’s degree in electrical and electronics engineering from Kerala University, Thiruvananthapuram, India, in 1995, and the master’s degree in computer and information sciences and the Ph.D. degree both from the Cochin University of Science and Technology, Kochi, India, in 1997 and 2009, respectively.

She worked as an Assistant Professor with the Department of Electrical and Electronics Engineering, Government Engineering College, Thrissur, India, from 1998 to 2009, where she is currently a Professor with the Department of Electrical Engineering. Her current research interests include power system operation and control, smart grid, Advanced Metering Infrastructure, microgrid control, etc.



T. P. Imthias Ahamed received the B.Tech. degree in electrical engineering from Kerala University, Thiruvananthapuram, India, in 1988, the M. Tech. degree in instrumentation and control from the National Institute of Technology Calicut, Kozhikode, India, in 1991, and the Ph.D. degree from the Indian Institute of Science, Bangalore, India, in 2002.

He has 25 years of teaching and research experience, which include three years in Kind Saud University, Riyadh, Saudi Arabia and two years in Dhofar University, Salalah, Oman. He is currently a Professor with the Department of Electrical and Electronics Engineering, Thangal Kunju Musaliar College of Engineering, Kollam, Kerala. His current research include reinforcement learning, demand response, neural networks, and power system scheduling and control.