Received: 12 August 2018 | Revised: 20 December 2019 | Accepted: 6 February 2020

DOI: 10.4218/etrij.2018-0446

ORIGINAL ARTICLE

ETRI Journal WILEY

Adaptive algorithm for optimal real-time pricing in cognitive radio enabled smart grid network

Deepa Das[1] | **Deepak Kumar Rout**[2]

[1]Department of Electrical Engineering, Government College of Engineering, Kalahandi, Odisha, India

[2]School of Electronics Engineering, KIIT University, Bhubaneswar, Odisha, India

Correspondence
Deepa Das, Department of Electrical Engineering, Government College of Engineering, Kalahandi, Odisha, India.
Email: deepadas.ctc@gmail.com

Integration of multiple communication technologies in a smart grid (SG) enables employing cognitive radio (CR) technology for improving reliability and security with low latency by adaptively and effectively allocating spectral resources. The versatile features of the CR enable the smart meter to select either the unlicensed or the licensed band for transmitting data to the utility company, thus reducing communication outage. Demand response management is regarded as the control unit of the SG that balances the load by regulating the real-time price that benefits both the utility company and consumers. In this study, joint allocation of the transmission power to the smart meter and consumer's demand is formulated as a two stage multi-armed bandit game in which the players select their optimal strategies noncooperatively without having any prior information about the media. Furthermore, based on historical rewards of the player, a real-time pricing adaptation method is proposed. The latter is validated through numerical results.

KEYWORDS
cognitive radio, demand response management, real-time pricing, smart grid, two-stage multi-armed bandit game

1 | INTRODUCTION

The traditional power grid is fully upgraded in its power generation, transmission, and distribution by incorporating adaptive control, embedded sensing, wireless communication networking, pervasive computing, and advanced two-way communication and networking between the power plants and the end users to improve its efficiency, reliability, stability, security, and sustainability [1,2]. This evidently leads to the power system of the next generation, called the smart grid (SG). A smart meter acts as the gateway that collects the energy consumption usages of the appliances at intervals and communicates them to the utility company for monitoring. This enables the utility company to use these data to update its electricity generation and determine a price that benefits itself and the consumers. The energy consumption scheduling of the appliances is managed by demand response management (DRM). The real-time pricing model used by the utility company encourages efficient and economic power utilization and distribution. However, balancing energy demand and load is also required to minimize the energy consumption cost and peak hour generation.

With the transformation of the traditional electric grid to the SG, the number of smart meters has increased substantially. This requires more frequency bands to accommodate such explosive growth of the data set from the smart meter, and this is one of the major research challenges in the SG. The cognitive radio (CR) technology in the SG provides an extensive solution to this problem. It enables efficient utilization of the limited spectrum, thus meeting the demands for more data rate and quality of service. In CR, the conflict between spectrum scarcity issues and underutilization of the spectrum band is solved by allowing the secondary users (SUs) to use the licensed band opportunistically as long as they do not cause any interference to the primary user (PU) [3–5]. This feature of CR makes it capable of solving this spectrum scarcity issues in multidisciplinary application domains such as machine to machine communication [6], vehicular networking [7], and satellite communication [8]. With the integration of CR technology in the home area network (HAN) of a SG, the smart meter has the capability to perform spectrum sensing and channel switching on either the original unlicensed or licensed band for data transmission to the utility company. However, the extra power consumption of the smart meter during this data transmission burdens the consumers with a high electricity price.

In the literature, different techniques have been proposed to minimize the power consumption and price of the consumers in a SG. In [9], Chou and others demonstrated an experiment by installing a SG metering infrastructure in a typical three-story building. The SG big data analytic framework could efficiently collect and analyze the real-time power consumption data. Furthermore, a novel hybrid nature–inspired metaheuristic forecast system was designed to predict the future energy consumption and scheduling operation of the appliances. In [10], the energy consumption of a residential grid network was managed by minimizing the electricity cost and the operational delay of flexible demands. An upgrade-by-probability was used to minimize this delay in which a delay-tolerant demand was treated as delay-sensitive demand and participated in a queue with a probability. Furthermore, a distributed algorithm was used to optimize the problem and a neural network was employed to estimate the pricing for an unknown system information. In [11], a multi-objective evolutionary algorithm was employed to minimize the energy cost and waiting time of the home appliances. Furthermore, an energy management algorithm was proposed for the residential home to maintain load balancing. A threshold was assigned for managing the operation of appliances to maintain the energy usage. In [12], the interaction between the utility company and consumers was formulated as a two-step centralized game. The objective of the game was to minimize the peak-to-average ratio (PAR) by optimizing the process of the energy consumption schedule. In this game theoretic approach, consumers maximized their own utility function using the interior-point method, and the adjustment of energy price of the utility company was based on consumers and their scheduling status. In [13], the interaction between a single utility company and its many customers was formulated as a 1-leader, *N*-follower Stackelberg game. An iterative algorithm was proposed to derive the real-time price, optimal power generation from the power provider, and optimal demands of the customers. In [14], He and others proposed a methodology based on a recurrent neural network to obtain the optimal real-time price in a SG. Furthermore, the convergence of the proposed approach was proved by using the theory of differential inclusions and a Lyapunov-like method. In [15], the interactions between consumers were formulated as a non-cooperative game. The consumers managed their energy consumption making a trade-off between electricity cost and load curtailment cost. The proposed approach was analyzed based on consumers with heating ventilation air conditioning systems. In [16], a community-based cooperative energy consumption scheduling scheme was proposed in which a group of customers in a community cooperatively minimized their PAR of the energy demand at different time periods. In [17], A multi-objective energy consumption scheduling approach under certain constraints was proposed to solve the problem of the two conflicting objectives of consumers that aim to maximize their utilities while minimizing energy consumption. In the algorithm, the iterative process was based on a genetic algorithm, the matrix crossover and polynomial mutation were based on Tchebycheff decomposition, and the constraint was handled by the matrix encoding mechanism.

CR technology has been nominated as a suitable candidate in SG communication networks to facilitate efficient and reliable data transmission over the underutilized spectrum by adaptively changing the operational parameters according to the dynamic surrounding environment. This has led to numerous research challenges in this application domain of CR technology. In the literature, different techniques have been proposed to solve some of the major challenges. In [18], the authors proposed a channel selection mechanism based on the transmitted data (either delay sensitive or non-delay sensitive) between the smart meter and the control center. Reliable power allocation with a relay association algorithm for a heterogeneous neighborhood area network was discussed in [19]. The objective was to maintain the transmission reliability above a certain threshold while limiting the power consumption considering all possible interferences with the smart meters. In [20], a cognitive mobile network with small cells was considered in a smart grid environment. The electricity price decision, the problems of retailers, the energy efficient power allocation to the small cell base stations, and the interference in the cognitive network were formulated as a three-stage Stackelberg game. In [21], the authors discussed a SG network integrated with TV white space (TVWS) cognitive radio. A HAN gateway enabled with cognitive capability switched between an industrial, scientific, and medical band and a TVWS channel depending on their availability. The authors extended the work considering both static and mobile HAN in [22]. Cooperative spectrum sensing in a SG environment was discussed in [23]. In this paper, some of the consumers were equipped with renewable power generation facilities. The communication outage probability was considered in the calculation of electricity price. Furthermore, a novel algorithm was proposed to find the optimal cooperative consumer for improving the DRM performance. The impact of communication outage on the performance of DRM was also considered in [24]. In this paper, the optimal sensing time that maximized the net revenue under the constraint of probability of detection was found.

From the literature, we observe that electricity price minimization is the major concern from the consumer's perspective, which depends on two important parameters such as consumer's demand and power generation from the utility company. Furthermore, in a CR-enabled SG network, the smart meter consumes some extra power to transmit its data set to the control unit of the power supplier. This maximizes the overall power demands from the consumers' side, and thus, enforces the power provider to supply more power. In consequence, the power supplier needs to raise the electricity price to profit itself. To solve this problem, this work addresses a cost-effective DRM scheme consisting of one utility company and multiple consumers. The consumers adjust their demands and transmission power to maximize their profit at the declared price from the power provider. Similarly, the power provider updates its electricity price and generation to ensure its profit and to suit all the consumers having different demands and transmission power. The main contributions of this study are:

- A time-varying real-time price-based DRM is proposed to manage the electricity trading process between the utility company and the consumers, so that it guarantees profit maximization to both the utility company and the consumers.
- The assignment of transmission power and power demands to the consumers is formulated as a multi-armed bandit (MAB) game under the constraints of probability of detection, interference to the primary receiver (PR), and minimum achievable throughput. Then, the electricity prices at different time slots are derived using the Hungarian method.
- Finally, the performance of our proposed work is presented and analyzed based on different network parameters.

The rest of this paper is organized as follows. The impact of communication outage on the DRM performance is derived in Section 2. Section 3 describes the optimization problem of both the power provider and consumers. Our proposed solution approach is discussed in Section 4. In Section 5, the performance of our proposed scheme is evaluated, and finally the paper is concluded in Section 6.

2 | SYSTEM MODEL

2.1 | Demand response management

The system model illustrates the bidirectional flow of information between the power provider and M consumers as shown in Figure 1. Each consumer is equipped with smart meter, which acts as a gateway between the power provider and consumers for distributing the electricity supply to all the appliances and reports the energy requests to the power provider.

Furthermore, the smart meter is connected to the CR node, which performs spectrum sensing on the licensed band before the data transmission. We divide the observation day duration T into L time slots, $\sum_{l=1}^{L} t_l = T$, that is, $t_l = \{t_1, t_2, t_3, \ldots, t_L\}$.

Let s_{t_l} be the time dependent power supply of the provider. Then, the cost function $C(s_{t_l})$ indicates the expenses of power supply for the t_l slot, which is strictly increasing and convex [24], that is,

$$C\left(s_{t_l}\right) = \frac{a_{t_l}}{2} s_{t_l}^2 + b_{t_l} s_{t_l} + c_{t_l}, \tag{1}$$

where a_{t_l}, b_{t_l}, and c_{t_l} are constants. Let p_{r,t_l} be the electricity price declared by the power provider at slot t_l. Then, the profit obtained by the power provider with power supply s_{t_l} during that time duration is given by $p_{\mathrm{r},t_l} s_{t_l} - C(s_{t_l})$. Hence, the goal of the provider is to maximize its own profit U_{T,t_l}, that is,

$$\max_{s_{t_l}} p_{\mathrm{r},t_l} s_{t_l} - C\left(s_{t_l}\right) \text{ s.t. } s_{t_l} \geq \sum_{m=1}^{M} d_{m,t_l} \tag{2}$$

where d_{m,t_l} is the power demand of the mth consumer at slot t_l. The constraint in (2) ensures that the power procured by all the consumers M at each time slot t_l must not exceed the maximum generation capacity of the power supply system. In general, d_{m,t_l} depends on the consumer type and is generated within the range $[d_{m,t_l}^{\min}, d_{m,t_l}^{\max}]$. $d_{m,t_l}^{\min}$ and $d_{m,t_l}^{\max}$ are the minimum and maximum demand of the mth consumer at slot t_l. For the power demand d_{m,t_l}, $G_{m,t_l}\left(d_{m,t_l}, w_{m,}\right)$ represents the mth consumer's

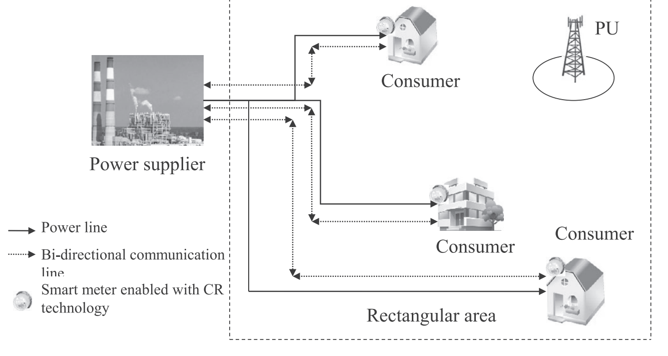

FIGURE 1 System model illustrating distribution of users

level of satisfaction function, which is a nondecreasing and concave function. Hence, the profit of the mth consumer is $R_{Wm,t_l} = G_{m,t_l}\left(d_{m,t_l}, w_m\right) - p_{r,t_l} d_{m,t_l}$. Therefore, the goal of the consumer is to maximize its profit, that is,

$$\max_{d_{m,t_l}} \; G_{m,t_l}\left(d_{m,t_l}, w_m\right) - p_{r,t_l} d_{m,t_l} \tag{3}$$

where

$$G_{m,t_l}\left(d_{m,t_l}, w_m\right) = \begin{cases} w_m d_{m,t_l} - \frac{\theta_m}{2} d_{m,t_l}^2 & 0 \le d_{m,t_l} \le \frac{w_m}{\theta_m} \\ \frac{w_m^2}{2} & d_{m,t_l} \ge \frac{w_m}{\theta_m} \end{cases}$$

where w_m is a parameter that characterizes the behavior of the mth consumer. The consumer always intends to consume more power, up to its saturation level. Furthermore, their satisfaction level gradually becomes saturated. θ_m is a constant. A higher value of θ_m requires lower energy consumption of the mth consumer to reach the saturation.

2.2 | Cognitive radio–enabled smart grid

The overwhelming demands for more spectrum to transmit bulk smart meter data require the integration of the CR technology into the SG network. The cognitive capability and reconfigurability features of CR help in selecting the best unutilized spectrum band from the temporally and spatially varying radio environment, which probably improves the communication quality and reliability in the SG.

In a HAN, the smart meter with cognitive features enables channel switching between the unlicensed and licensed bands for information exchange between the smart meter and the utility company. The smart meter performs spectrum sensing before opting for the suitable channel for data transmission. Let P_{H1} and P_{H0} be the probability of PU's presence and absence on the licensed band, respectively. Channel switching from the unlicensed to the licensed band is possible in two cases: when the channel is correctly detected to be idle, which occurs with probability $P_{H0}\left(1 - P_{f,m}\right)$ and when the PU is in operation but the channel is identified as vacant, which occurs with probability $P_{H1} P_{md,m}$. Here, $P_{f,m}$ and $P_{md,m}$ represent the false alarm and miss detection probability for the mth smart meter, respectively. Thus, probability of channel switching of the mth smart meter is given as follows:

$$C_{sw,t_l}(m) = P_{H0}\left(1 - P_{f,m}\right) + P_{H1} P_{md,m}. \tag{4}$$

Let O_U and O_L represent the outage of the unlicensed and licensed bands, respectively. The smart meter performs no switching operation from the unlicensed to the licensed band when the PU is in operation and no false alarm occurs. It occurs with probability $\left(1 - C_{sw,t_l}(m)\right) O_U$. Channel switching to the licensed band occurs with probability $C_{sw,t_l}(m)\, O_L$. Hence, the total channel outage probability is defined as follows:

$$\begin{aligned} O_{g,t_l}(m) &= \left(1 - C_{sw,t_l}(m)\right) O_U + C_{sw,t_l}(m)\, O_L \\ &= \left(O_L - O_U\right) C_{sw,t_l}(m) + O_U \end{aligned} \tag{5}$$

In this study, it is assumed that $O_L < O_U$ so as to give priority to the licensed band considering its good quality of transmission and reliability. If $O_{g,t_l}(m) < O_U$, the mth smart meter needs to perform spectrum sensing and channel switching.

According to the central limit theorem, the local test statistics is a Gaussian distribution for a large value of samples. Assuming that PU's signal is a binary phase shift keying signal and noise is real, $P_{d,m}$ and $P_{f,m}$ are represented as [25]

$$P_{d,m} = Q\left(\frac{\lambda - \sigma_{\eta m}^2\left(1 + \gamma_m\right)}{\sigma_{\eta m}^2 \sqrt{2\left(1 + 2\gamma_m\right)}} \sqrt{N}\right) \tag{6}$$

and

$$P_{f,m} = Q\left(\frac{\lambda - \sigma_{\eta m}^2}{\sigma_{\eta m}^2 \sqrt{2}} \sqrt{N}\right) \tag{7}$$

where $\sigma_{\eta m}^2$ is the noise variance of the mth smart meter and γ_m is the received signal-to-noise ratio (SNR) at the mth smart meter. N denotes the number of samples and is the product of sensing time t_s and sampling frequency f_s. λ represents the decision threshold. $P_{md,m}$ measures the interference to the incumbent PU, which must be kept to a minimum. Hence, $P_{md,m} \le \overline{P}_{md}$ or $P_{d,m} \ge \overline{P}_d$ or $P_{d,m} \ge \overline{P}_d$ where $\overline{P}_{md}$ and $\overline{P}_d$ are the upper and lower bounds of miss detection probability and probability of detection, respectively. Hence, threshold λ for $\overline{P}_d$ is calculated as follows:

$$\lambda = Q^{-1}\left(\overline{P}_d\right) \frac{\sigma_{\eta m}^2 \sqrt{2\left(1 + 2\gamma_m\right)}}{\sqrt{N}} + \sigma_{\eta m}^2\left(1 + \gamma_m\right). \tag{8}$$

Substituting (8) in (7), $P_{f,m}$ is modified in terms of $\overline{P}_d$ as follows:

$$\tilde{P}_{f,m} = Q\left(\frac{Q^{-1}\left(\overline{P}_d\right) \sigma_{\eta m}^2 \sqrt{2\left(1 + 2\gamma_m\right)} + \sigma_{\eta m}^2 \gamma_m \sqrt{N}}{\sigma_{\eta m}^2 \sqrt{2}}\right). \tag{9}$$

Substituting (9) and $\overline{P}_{\text{md}}$ in (4),$C_{\text{sw},t_l}(m)$ is evaluated and accordingly the total channel outage probability is evaluated for the mth smart meter at slot t_l.

3 | PROBLEM FORMULATION

In a CR integrated SG, the smart meter transmits its data on the licensed band under the interference power constraint, so that its transmission power does not cause any interference to the PR. Furthermore, throughput increases with the transmission power of the smart meter. Hence, transmission power of the mth consumer at slot t_l, that is, P_{tm,t_l} must be in the range $[P^{\min}_{\text{tm},t_l}, P^{\max}_{\text{tm},t_l}]$. $P^{\min}_{\text{tm},t_l}$ is decided such that the throughput of the mth consumer at slot t_l is $T_{m,t_l} \geq T_{\text{th}}$, where T_{th} is the minimum achievable throughput. Similarly, $P^{\max}_{\text{tm},t_l}$ is the maximum transmission power from the smart meter at slot t_l, which depends on the PR distance from the mth consumer and interference threshold I_{th}. I_{th} is the maximum interference power that the PR can tolerate. Hence, the power assigned to the mth consumer at slot t_l, P_{tm,t_l}, must comply with these two constraints, (10a) and (10b).

$$P^{\min}_{\text{tm},t_l} \geq \frac{\left(2^{T_{\text{th}}}-1\right) N_{\text{p}}}{\left|h_{\text{sd}}\right|^2} \tag{10a}$$

and

$$P^{\max}_{\text{tm},t_l} \leq \frac{I_{\text{th}}}{|h_{\text{pr}}|^2} \tag{10b}$$

where N_{p} is the noise power. h_{sd} and h_{pr} are the channel coefficients between the consumer and the supplier and between the consumer and the PR, respectively. The communication channel coefficients between the users are distance-dependent Rayleigh distributed and are derived as $h \sim \aleph(1,(1/d^{\alpha}))$, where d is the distance between the users and α is the path-loss exponent. Furthermore, the communication outage $O_{\text{g},t_l}(m)$ has greater impact in deriving R_{Wm,t_l} of the mth consumer. Hence, R_{Wm,t_l} and s_{t_l} are modified as follows:

$$\begin{aligned} R_{Wm,t_l} =& G_{m,t_l}\left(d_{m,t_l}, w_m\right) \\ & - p_{\text{r},t_l}\left(O_{\text{g},t_l}(m).d_{m,t_l} + C_{\text{sw},t_l}(m).P_{\text{tm},t_l} + P_{\text{s}}\right) \end{aligned} \tag{11}$$

and

$$s_{t_l} \geq \sum_{m=1}^{M}\left(1-O_{\text{g},t_l}(m)\right) d_{m,t_l} \tag{12}$$

where P_{s} is the sensing power. The impact of $O_{\text{g},t_l}(m)$ on the supplier is clearly observed from (12). It is shown that the power demand of each consumer d_{m,t_l} estimated at the control unit of the supplier is $\left(1-O_{\text{g},t_l}(m)\right) d_{m,t_l}$. If we compare (2) and (12), it is observed that power generation also reduces due to this communication outage, and hence, it is unable to meet the total power demand of the consumers. Owing to this reason, the supplier needs to consider the communication outage while deciding its exact power supply. In this two-stage 1-to-M player MAB game, the goal of the mth player is to maximize its own profit/reward function (11) under the constraints of (10a) and (10b). All M players play individually and update their demands and transmission powers with the price p_r declared by the power provider. Furthermore, (11) shows the existence of Nash equilibrium (NE) in this non-cooperative game theory, which draws to an optimal solution given as follows:

Lemma 1 *For a constant P_{tm,t_l}, a NE exists in the game $G = \left[M, d_{m,t_l}, R_{Wm,t_l}\right] \forall d_{m,t_l}$ if*

- *R_{Wm,t_l} is continuous and concave in d_{m,t_l}.*
- *d_{m,t_l} is non-empty, convex, and bounded.*

Proposition *A NE exists in this two-stage game.*

Proof The strategy d_{m,t_l} is defined as $d_{m,t_l}: d^{\min}_{m,t_l} \leq d_{m,t_l} \leq d^{\max}_{m,t_l}$.

For a constant P_{tm,t_l}, the first derivative of (11) yields

$$\frac{dR_{Wm,t_l}}{d_{m,t_l}\Big|_{P_{\text{tm},t_l}=\text{constant}}} = w_m - \theta_m d_{m,t_l} - p_{\text{r},t_l} O_{\text{g},t_l}(m). \tag{13}$$

If the Hessian matrix $H\left(R_{Wm,t_l}\right)$ is negative, then R_{Wm,t_l} is strictly concave in d_{m,t_l}. The second derivative of (11) with respect to d_{m,t_l} for a constant P_{tm,t_l} is obtained as follows:

$$\frac{d^2 R_{Wm,t_l}}{d^2_{m,t_l}\Big|_{P_{\text{tm},t_l}=\text{constant}}} = \begin{cases} -\theta_m, & t_l = t \\ 0 & t_l \neq t \end{cases}, \tag{14}$$

where t denotes any slot in the duration T. Hence, the diagonal elements of $H\left(R_{Wm,t_l}\right)$ are negative and the other elements are zero. It is proved that R_{Wm,t_l} is strictly concave in d_{m,t_l} showing the existence of a unique NE for obtaining the best response strategy at each step.

4 | SOLUTION APPROACH

4.1 | MAB-based two-stage game

In any learning algorithm, exploitation and exploration are equally important [26]. Exploitation finds the strategy that gives the best reward. Exploration allows the player to play with all the strategies to find their reward statistics. MAB is an extensive solution to model the trade-off between exploration and exploitation with the aim to obtain the best response. The action of the player is to select one arm among many arms at each step. The MAB is a strategic decision-making game in which each player selects the strategy/arm randomly in a sequence from the defined strategies of trials with a goal to maximize its own reward. However, the method makes a trade-off between exploration (try each arm to find the best reward) and exploitation (play the arm and forecast the best result by analyzing the present and history data). With the subsequent steps, the players achieve the best reward giving strategy.

As the consumers/players have their own rights of choosing their strategies, exchange of information between the players is not required in this method. Rather, the player analyzes the historical response of the strategies before achieving any decision. The MAB has been applied to the CR domain for resource allocation in [27–29]. The aforementioned MAB-based approaches effectively selected the best response with poor a priori information. In this study, we proposed a two-stage MAB game in which the strategy is the joint achievement of demands and the corresponding transmission power assigned to each player. Thus, the m th player declares two feasible strategies P^{x}_{tm,t_l} and d^{d}_{m,t_l} of length X and D, respectively for slot t_l as follows:

$$P^{x}_{\mathrm{tm},t_l} = \left\{ P^{1}_{\mathrm{tm},t_l}, P^{2}_{\mathrm{tm},t_l}\ , \ldots, P^{X}_{\mathrm{tm},t_l} \right\} \tag{15}$$

and

$$d^{d}_{m,t_l} = \left\{ d^{1}_{m,t_l}, d^{2}_{m,t_l}, \ldots, d^{D}_{m,t_l} \right\}. \tag{16}$$

In this work, the reward of each player is the utility function defined in (11). It is very clearly observed that the mth player has to select suitable strategies from (15) and (16) with an appropriate combination to achieve the best response. Hence, the players use this learning approach between these two feasible strategies, which is thus called the two-stage MAB game. The UCB1 algorithm is often used to solve the MAB problem that selects the strategy to play based on the past strategies that were played [30]. It is based on the rewards associated with each arm and selects the arm with the highest/best reward. In the subsequent steps, the reward mean gradually increases until the best arm is selected. According to the UCB1 algorithm, the trade-off between exploitation and exploration is obtained as follows:.

$$R_s = \arg\max_{s \in S} \left(\hat{\mu}_s + \sqrt{\frac{a \ln Y}{m_s}} \right) \tag{17}$$

where $\hat{\mu}_s$ is the sample mean of reward of arm S up to the current state, m_s is the number of times strategy s is played up to the current time slot, Y is the number of times strategy s is played by the mth player, and a is the constant that controls exploitation and exploration, where a small value of a indicates small exploitation and large exploration, and vice versa. Therefore, at each step, the reward of each strategy is updated, and finally, the strategy with the highest mean reward is selected.

4.2 | Algorithm description

In this study, the real-time pricing method is considered observing the demands in different time slots of a day. The pseudocode for obtaining optimal price, generation from the power provider, consumer's demand, and transmission power is presented in Algorithm 1. We consider that the total duration of a day T = 24 h is divided into four distinct slots t_1 = 6 AM to 9 AM, t_2 = 9 AM to 5 PM, t_3 = 5 PM to 10 PM, and t_4 = 10 PM to 6 AM. Furthermore, we consider three types of consumers $M = 3$. Consumer 1 and consumer 3 belong to residential areas, but consumer 3 is a working person staying outside home during slot t_2. Consumer 2 is assumed to be an office. Hence, their demands vary in different time slots depending on the appliances used during that period. For each slot, the power provider declares its generation depending on consumers' target demands as per $s_{t_l} = \sum_{m=1}^{M} d_{m,t_l}$. The smart meter needs to report the power supplier regarding the usages of the appliances on the licensed band; hence, it has to select the transmission power P_{tm,t_l} satisfying the constraints (10a) and (10b). Therefore, the transmission power levels of the consumers are different for t_1, t_2, t_3, and t_4.

In Algorithm 1, the demand of the consumer d_{m,t_l} and transmission power from the smart meter P_{tm,t_l} are considered as two strategies. Then, d_{m,t_l} of size D and P_{tm,t_l} of size X are randomly generated. In this two-stage MAB game, first, the demand of each mth consumer acts as a player that tries the second strategy, that is, its assigned transmission power, to maximize the reward (11). Each player randomly picks the strategy, that is, not played, and calculates its reward function. When all the players have played all the strategies, P_{tm,t_l} is calculated using the UCB1 algorithm. The corresponding player that selects P_{tm,t_l} is selected as d_{m,t_l} of the mth consumer. Furthermore, (13) and (14) show the existence of NE for obtaining the best response from (11). With this, d_{m,t_l} and its corresponding reward R'_{Wm,t_l} are obtained from Step 5. The power supplier declares new generation considering all the consumers' demands obtained from Step 6 and communication outage $O_{g,t_l}(m)$. Then, p_{r,t_l} is updated for slot t_l by comparing the current and last generation. As per our proposed Algorithm 1, the power supplier generates p_{r,t_l} of size A, which ranges from $p^{\min}_{r,t_l}$ to $p^{\max}_{r,t_l}$. The minimum and maximum price $p^{\min}_{r,t_l}$ and $p^{\max}_{r,t_l}$, respectively, are decided taking ±50 % of the p_{r,t_l}. Then, the rewards of all the consumers are calculated for each price keeping their P_{tm,t_l} and d_{m,t_l} unchanged. The matrix R'_{W,t_l} is solved by using the Hungarian method for obtaining the best price p^1_{r,t_l}, p^2_{r,t_l}, and p^3_{r,t_l} for consumer 1, consumer 2, and consumer 3, respectively. From these three prices, the power provider selects the price that maximizes its profit (2).

Algorithm 1. Two-stage game-based DRM

Step 1: Initialization and Declaration

Step 2: for $t \le T$

Initialization

for $m = 1{:}M$

Generate P^x_{tm,t_l} and d^d_{m,t_l} as per (15) and (16), respectively;

Calculate $C_{\mathrm{sw},t_l}(m)$ and $O_{\mathrm{g},t_l}(m)$ as per (4) and (5), respectively;

Initialization

while $n \le X + 2$

$n = n + 1$;

for $d = 1{:}D$

the mth player randomly picks the demand d_{m,t_l} from d^d_{m,t_l}, that is, not played;

if $n \le X$

the mth player randomly picks the transmission power P_{tm,t_l} from P^x_{tm,t_l}, that is, not played;

Evaluate R_{Wm,t_l};

$y(n,d) = \frac{y(n,d)q(n,d) + R_{Wm,t_l}}{q(n,d)+1}$;

else

$P_{\mathrm{tm},t_l} = \underset{x \in X}{\operatorname{argmax}}\, y(n,d) + \sqrt{\frac{a\ln(n)}{q(n,d)}}$;

end if

end for

end while

Step 3: Find $P^d_{\mathrm{tm},t_l} = \underset{p^d_{\mathrm{tm},t_l}|d \in D}{\operatorname{argmax}} R_{Wm,t_l}$;

Step 4: Evaluate corresponding d_{m,t_l} and R_{Wm,t_l};

Step 5: Find $d_{m,t_l} = \underset{d \in D}{\operatorname{argmax}} R_{Wm,t_l}$ and the corresponding R'_{Wm,t_l};

Step 6: if $R_{Wm,t_l} \ge R'_{Wm,t_l}$;

d_{m,t_l} is according to Step 4;

else

d_{m,t_l} is according to Step 5;

end if

end for

Step 7: $s'_{t_l} = \sum_{m=1}^{M} \left(1 - O_{\mathrm{g},t_l}(m)\right) d_{m,t_l}$;

Step 8: Power supplier declares new p_{r,t_l} based on s_{t_l} and s'_{t_l};

Step 9: $p^a_{\mathrm{r},t_l} = \left[p^1_{\mathrm{r},t_l}, p^2_{\mathrm{r},t_l}, \ldots, p^A_{\mathrm{r},t_l}\right]$ s.t $p^{\min}_{\mathrm{r},t_l} \le p^a_{\mathrm{r},t_l} \le p^{\max}_{\mathrm{r},t_l}$;

Thus, the real-time prices at different time slots are updated based on the target demands, transmission powers of the consumers, and outage on the licensed band. Furthermore, the outage depends on the distance of the consumer from the primary transmitter and the wireless channel media. Similarly, the transmission power from the smart meter is decided based on the distance of the mth consumer from the PR and the power provider. Therefore, the price at each time slot t_l is decided carefully considering all these factors from the consumer's aspect so that it benefits both the consumers and power supplier.

5 | NUMERICAL RESULTS

We consider a network consisting of three consumers and a primary transmitter which are distributed in a rectangular area 1 km in width. The PR is present within a circular radius of 50 m around the primary transmitter. The power provider is present approximately 2 km away from the consumers. Initially, we set the electricity price $p_\mathrm{r} = 1$ (cent/kWh). Then, it is updated as per the consumer's demands and supplier's generations. Different values of a_t are considered for different time slots. We set $a_t = 0.02$ for t_1, t_2, and t_3, $a_t = 0.01$ for t_4, $b_t = 0.2$, and $c_t = 0$. For consumer 1, consumer 2, and consumer 3, w_m is taken as 5, 6, and 5, respectively. We set $a = 0.5$ and $\theta_m = 0.1$ for all the consumers. The target demands of consumer 1, consumer 2, and consumer 3 for different time slots are listed in Table 1. The minimum and maximum value of the consumer's demand is decided by taking ±50% of the target demand. The rest of the common parameters that are used to obtain the results are summarized in Table 2.

Figure 2 illustrates the variation of price over the time slots that are obtained by applying the UCB1 algorithm, NE, and Algorithm 1. In the UCB1 algorithm, d_{m,t_l} and P_{tm,t_l} are obtained from Algorithm 1 without using Step 5 and Step 6. In NE, d_{m,t_l} is obtained by following only Step 5 and Step 6 for the mth consumer. We consider three different types of consumers located at different distances from the primary transmitter, PR, and power supplier. We assume that the received SNRs are different. These are set as −15 dB, −25 dB, and −30 dB for consumer 1, consumer 2, and consumer 3, respectively. The same SNR conditions are also considered in all the figures except Figures 4 and 7. It is observed from Figure 2 that even though the power supplier declares a constant price $p_\mathrm{r} = 1$ for each slot, it varies as per the demands from the consumers at different time slots. Furthermore, the updated price obtained from Algorithm 1 is less than that from UCB1 and NE algorithms, which is more beneficial to the consumers.

Figure 3 shows the variation of consumers' demands over different time slots. It is observed that the consumers' demands calculated from Algorithm 1 are approximately equal to the target demands. Figure 4 illustrates the impact of sensing time over communication outage

TABLE 1 Consumer's demands at different time slots

Time slots	Duration	Consumer 1 demand (kWh)	Consumer 2 demand (kWh)	Consumer 3 demand (kWh)
t_1	6 AM–9 AM	5	10	15
t_2	9 AM–5 PM	15	50	2
t_3	5 PM–10 PM	7	20	9
t_4	10 PM–6 AM	3	8	3

TABLE 2 System parameters

Simulation parameters	Values
f_s	6 MHz
t_s	3 ms
α	3
$\overline{P}_d$	0.9
N_p	−90 dBm
T_{th}	10 bits/s/Hz
P_s	−20 dBm
I_{th}	−50 dBm
P_{H0}	0.5
P_{H1}	0.5
O_U	0.3
O_L	0.2

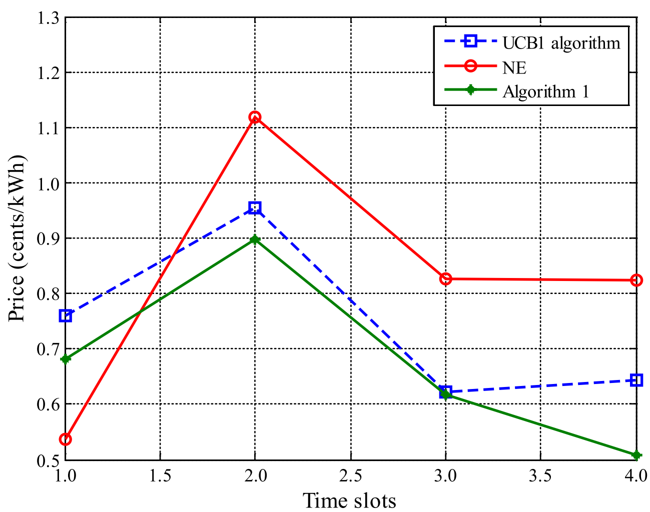

FIGURE 2 Variation of price over different time slots

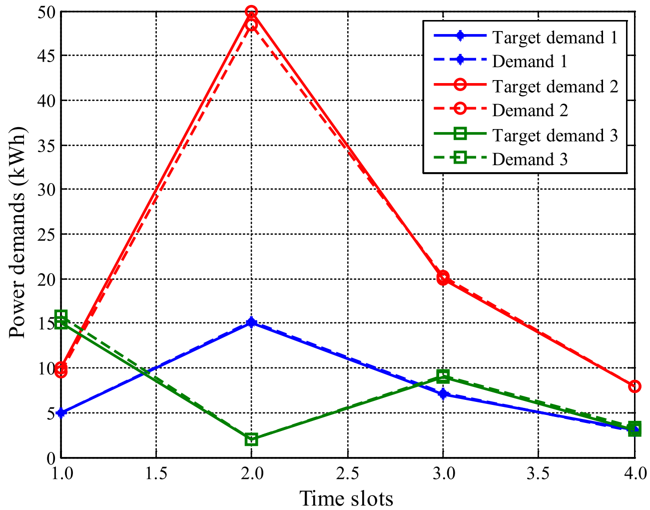

FIGURE 3 Variation of consumers' power demands over time slots

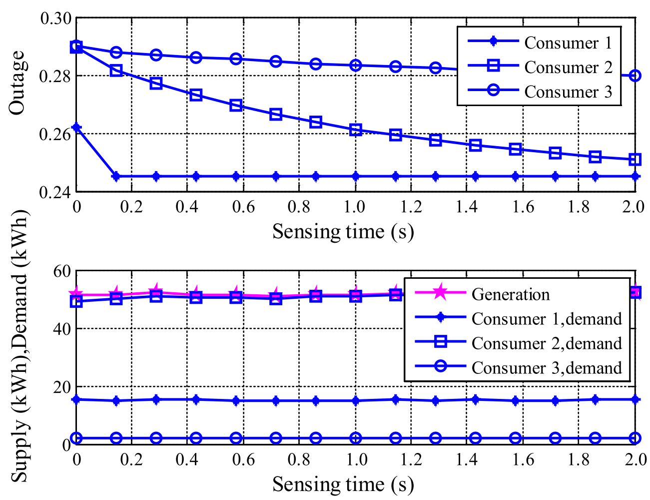

FIGURE 4 Impact of sensing time t_s on outage, power supply, and consumers' demand for time slot 2

on the licensed band, power generation, and consumers' power demands for time slot 2. The upper figure shows the decrease in communication outage with respect to sensing time. This is because the Q-function is a decreasing function; hence, $\tilde{P}_{f,m}$ decreases with the increase in t_s for a constant $\overline{P}_d$ as observed from (9). It means that the spectrum sensing accuracy increases with sensing time. Substituting the value of (9) in (4), it is found that $C_{sw,t_l}(m)$ increases with an increase in t_s. When the spectrum sensing is more accurate, then it is more likely that data transmission occurs on the licensed band. Hence, with an increase in sensing time and decrease in false alarm, total outage $O_{g,t_l}(m)$ decreases. It can be concluded that a longer sensing time ensures accurate spectrum sensing, thereby reducing the outage on the licensed band. We set the received SNRs as −15 dB, −30 dB, and −35 dB for consumer 1, consumer 2, and consumer 3, respectively. The bottom figure of Figure 4 shows the impact

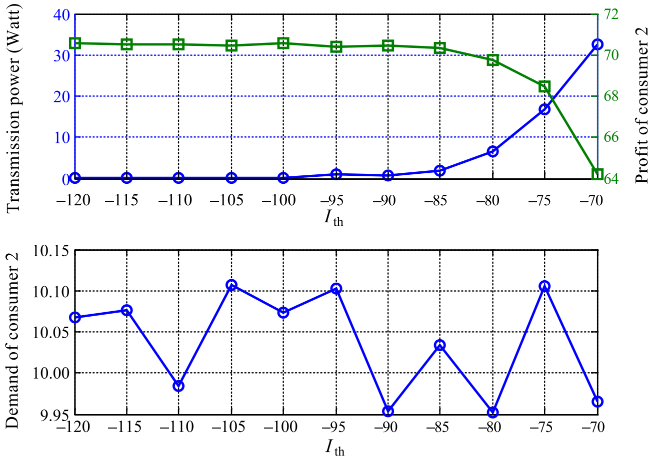

FIGURE 5 Effect of I_{th} on transmission power, profit, and power demand of consumer 2 for time slot 1

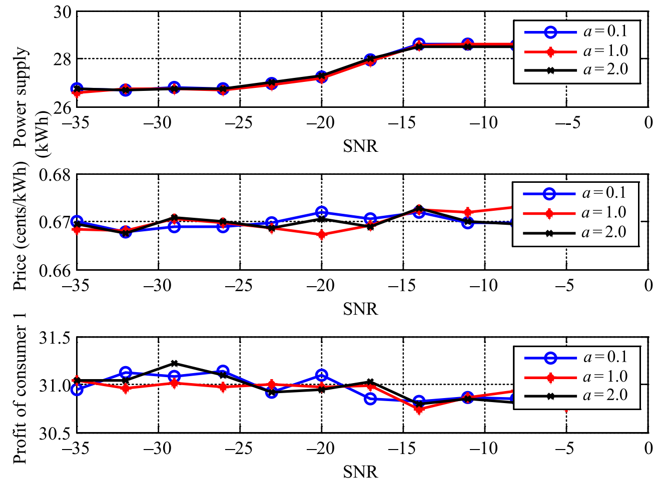

FIGURE 7 Variation of power supply, price, and profit of consumer 1 over SNR

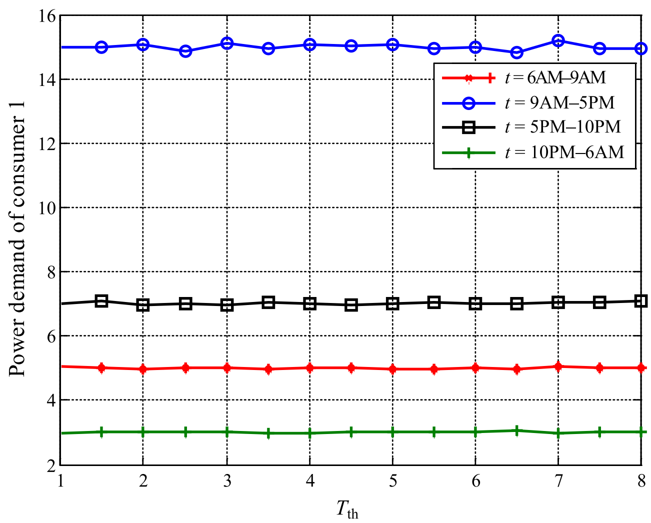

FIGURE 6 Effect of T_{th} on power demand of consumer 1

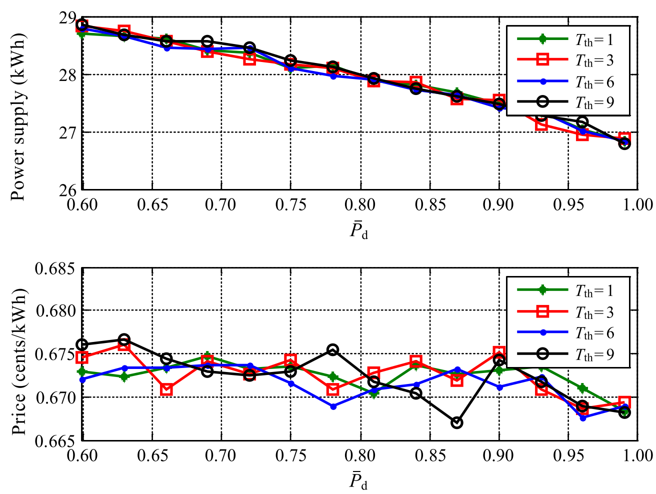

FIGURE 8 Variation of average power supply and price over target detection probability $\bar{P}_d$

of outage on the two-way communication in the SG. In general, the consumers situated in different locations suffer from different outages while sending their smart meter data. Therefore, the power demand of the mth consumer estimated at the control unit of the power supplier is $\left(1 - O_{g,t_l}(m)\right) d_{m,t_l}$. Hence, the power supply estimated at the supplier is $s_{t_l} = \sum_{m=1}^{M} \left(1 - O_{g,t_l}(m)\right) d_{m,t_l}$ instead of $s_{t_l} = \sum_{m=1}^{M} d_{m,t_l}$. Furthermore, according to (11), the satisfaction gain function of the consumer $G_{m,t_l}\left(d_{m,t_l}, w_m\right)$ is designed as ideal without considering the communication outage, but the expenses consider the communication outage. This figure shows the power generation and consumers' demand after applying Algorithm 1. d_{m,t_l} and P_{tm,t_l} of the consumers are obtained by applying the MAB two-stage game from Step 1 to Step 6 of Algorithm 1. Then, the power generation is estimated in Step 7 considering the communication outage. Therefore, the total generation is not the sum of the three consumers' demands. It is rather less than the total demands. To solve this problem, the power supplier must know the information on channel state of all the consumers before deciding s_{t_l}, which is one of the major research challenges in a CR-enabled SG network. It is also observed that power generation slightly increases with sensing time as per (12). However, the power demands from the consumers remain unchanged with the change in sensing time.

Figure 5 demonstrates the impact of I_{th} on the different outcomes of consumer 2, such as its data transmission power, reward (11), and power demand for time slot 1. It is obvious that P_{tm,t_l}^{max} increases with the increase in I_{th}, and thus, P_{tm,t_l} increases. Therefore, profit to consumer 2 decreases as per (11), but I_{th} has almost no impact on the power demand of consumer 2. The target demand of consumer 2 is

10 kWh; therefore, the actual power demand varies around this value.

Figure 6 illustrates the variation of power demand of consumer 1 over different values of T_{th}. It is observed that the power demand remains almost unchanged for different T_{th} values. From (10a), it is observed that $P^{\min}_{tm,t_l}$ increases with an increase in T_{th}. However, the selection of d_{m,t_l} from Algorithm 1 is not affected significantly with the increase in T_{th}.

Figure 7 illustrates the variation of power generation, price, and profit of consumer 1 for different values of the received SNR. It is assumed that all the consumers receive the same SNR. As the received SNR by each consumer increases, the communication outage decreases on the licensed band. Therefore, power generation from the utility company increases as per (12). As the communication outage decreases, the probability of switching to the licensed band increases. In general, the profit of consumer 1 decreases, but it does not change noticeably over SNR. Furthermore, it is observed that the impact of a on the different outcomes is negligible.

Figure 8 demonstrates the impact of target detection probability $\overline{P}_{\text{d}}$ on the average power generation and price. It is observed that power generation decreases with the increase in $\overline{P}_{\text{d}}$. This is because as $\overline{P}_{\text{d}}$ increases, $\tilde{P}_{\text{f},m}$ increases. Hence, $C_{\text{sw},t_l}(m)$ decreases as per (4). Thus, a higher probability of detection indicates a higher probability of PU's presence on the licensed band, which reduces the chances of channel switching from the unlicensed to the licensed band. This leads to an increase in communication outage, and hence, as per (12), power generation decreases with an increase in $\overline{P}_{\text{d}}$. It is more likely that the smart meter transmits its data on the unlicensed band when the PU is present on the licensed band. Price also decreases with the increase in $\overline{P}_{\text{d}}$, but the rate of decrease does not change significantly.

6 | CONCLUSIONS

This study discusses a solution to the problem of resource allocation to consumers in a CR-enabled SG network considering the profits of both the consumers and power provider. The problem of joint optimization of transmission power and consumer demand is formulated as an MAB-based two-stage game under the constraint of interference power to the legitimate user on the licensed band. Furthermore, the transmission outage is considered while obtaining the rewards of the consumers. Then, the optimal real-time price for all the consumers is declared by using the Hungarian method, which maximizes the profit of both the consumers and power provider. The performance of our proposed approach is analyzed through numerical results, and the effect of different system parameters is also studied. Furthermore, it is verified that the power supplier must consider the communication outage before deciding the actual demands from the consumers, which requires the knowledge of information on channel state of all the consumers. This is a major challenge in a CR-enabled SG network, which would be our focus in future research.

ORCID

Deepa Das https://orcid.org/0000-0002-0997-3082

REFERENCES

1. M. Erol-Kantarci and H. T. Mouftah, *Energy-efficient information and communication infrastructures in the smart grid: A survey on interactions and open issues*, IEEE Commun. Surv. Tuts. **17** (2015), no. 1, 179–197.
2. A. A. Khan, M. H. Rehmani, and M. Reisslein, *Cognitive radio for smart grids: survey of architectures, spectrum sensing mechanisms, and networking protocols*, IEEE Commun. Surv. Tuts. **18** (2015), no. 1, 860–898.
3. A. O. Bicen, O. B. Akan, and V. C. Gungor, *Spectrum-aware and cognitive sensor networks for smart grid applications*, IEEE Commun. Mag. **50** (2012), no. 5, 158–165.
4. T. N. Le, W. L. Chin, and H. H. Chen, *Standardization and security for smart grid communications based on cognitive radio technologies - A comprehensive survey*, IEEE Commun. Surv. Tuts. **19** (2017), no. 1, 423–445.
5. D. Das and S. Das, *Optimal resource allocation for soft decision fusion-based cooperative spectrum sensing in cognitive radio networks*, Comput. Electr. Eng. **52** (2016), 362–378.
6. Z. Zhou et al., *Energy-efficient resource allocation for energy harvesting-based cognitive machine-to-machine communications*, IEEE Trans. Cognitive Commun. Netw. **5** (2019), no. 3, 595–607.
7. K. Baraka et al., *An infrastructure aided cooperative spectrum sen sing scheme for vehicular ad hoc networks*, Ad Hoc Netw. **25** (2015), 197–212.
8. M. Hoyhtya et al., *Application of cognitive radio techniques to satellite communication*, in Proc. IEEE Int. Symp. Dynamic Spectrum Access Netw. (Bellevue, WA, USA), 2012, pp. 540–551.
9. J. S. Chou and N. T. Ngo, *Smart grid data analytics framework for increasing energy savings in residential buildings*, Autom. Constr. **72** (2016), 247–257.
10. Y. Liu et al., *Queuing-Based Energy consumption management for heterogeneous residential demands in smart grid*, IEEE Trans. Smart Grid **7** (2016), no. 3, 1650–1659.
11. K. Muralitharan, R. Sakthivel, and Y. Shi, *Multiobjective optimization technique for demand side management with load balancing approach in smart grid*, Neurocomput. **177** (2016), 110–119.
12. Z. M. Fadlullah et al., *GTES: An optimized game-theoretic demand-side management scheme for smart grid*, IEEE Syst. J. **8** (2014), no. 2, 588–597.
13. M. Yu and S. H. Hong, *Supply-demand balancing for power management in smart grid: A Stackelberg game approach*, Appl. Energy **164** (2016), 702–710.
14. X. He et al., *A recurrent neural network for optimal real-time price in smart grid*, Neurocomput. **149** (2015), 608–612.
15. K. Ma, G. Q. Hu, and C. J. Spanos, *Distributed energy consumption control via real-time pricing feedback in smart grid*, IEEE Trans. Contr. Syst. Technol. **22** (2014), no. 5, 1907–1914.

16. S. Bera, S. Misra, and D. Chatterjee, *C2C: Community-based cooperative energy consumption in smart grid*, IEEE Trans. Smart Grid **9** (2017), no. 5, 4262–4269.
17. H. Lu et al., *Multi-objective energy consumption scheduling in smart grid based on Tchebycheff decomposition*, IEEE Trans. Smart Grid **6** (2015), no. 6, 2869–2883.
18. S. Althunibat, Q. Wang, and F. Granelli, *Flexible channel selection mechanism for cognitive radio based last mile smart grid communications*, Ad Hoc Netw. **41** (2016), 47–56.
19. M. Haris et al., *Power allocation for reliable smart grid communication employing neighborhood area networks*, in Proc. IEEE Global Commun. Conf. (Abu Dhabi, United Arab Emirates), Dec. 2018, pp. 1–6.
20. S. Bu and F. R. Yu, *Green cognitive mobile networks with small cells for multimedia communications in the smart grid environment*, IEEE Trans. Veh. Technol. **63** (2014), no. 5, 2115–2126.
21. A. S. Cacciapuoti et al., *Enabling smart grid via TV white space cognitive radio*, in Proc. IEEE Int. Conf. Commun. Workshop (London, UK), June 2015, pp. 568–572.
22. A. S. Cacciapuoti et al., *Mobile smart grids: Exploiting the TV white space in urban scenarios*, IEEE Access **4** (2016), 7199–7211.
23. K. Zhang et al., *Cooperation for optimal demand response in cognitive radio enabled smart grid*, in Proc. IEEE Int. Conf. Commun. (Kuala Lumpur, Malaysia), May 2016, pp. 1–6.
24. R. Deng et al., *Sensing-performance tradeoff in cognitive radio enabled smart grid*, IEEE Trans. Smart Grid **4** (2013), no. 1, 302–310.
25. Y.-C. Liang et al., *Sensing-throughput tradeoff for cognitive radio networks*, Wireless Commun. IEEE Trans. **7** (2008), no. 4, 1326–1337.
26. H. Liu, K. Liu, and Q. Zhao. *Learning and sharing in a changing world: Non-Bayesian restless bandit with multiple players*, in Proc. Inf. Theory Appl. Work. (Lajolla, CA, USA), Feb. 2011, pp. 240–246.
27. S. Maghsudi and S. Stańczak, *Joint channel selection and power control in infrastructureless wireless networks: a multiplayer multiarmed bandit framework*, IEEE Trans. Veh. Technol. **64** (2015), no. 10, 4565–4578.
28. N. Modi, P. Mary, and C. Moy, *QoS driven channel selection algorithm for cognitive radio network: Multi-user multi-armed bandit approach*, IEEE Trans. Cogn. Commun. Netw. **3** (2017), no. 1, 49–66.
29. W. Wang et al., *Learning for robust routing based on stochastic game in cognitive radio networks*, IEEE Trans. Commun. **66** (2018), no. 6, 2588–2602.
30. P. Auer, N. Cesa-Bianchi, and P. Fischer, *Finite-time analysis of the multiarmed bandit problem*, Mach. Learn. **47** (2002), 235–256.

AUTHOR BIOGRAPHIES

Deepa Das is an assistant professor at the Department of Electrical Engineering of Government College of Engineering, Kalahandi, India. She received her BE in electronics and telecommunication engineering in 2005 from Utkal University, Odisha, India, her MTech degree in communication engineering in 2008 from KIIT University, Bhubaneswar, India, and her PhD degree in 2017 from the National Institute of Technology, Rourkela, India. She is a reviewer in many international journals. Her research interests include cognitive radio networks and its application in different domains, physical layer security, and heterogeneous networks.

Deepak Kumar Rout received his BTech degree in electronics and telecommunication engineering from Biju Patnaik University of Technology, Rourkela, India in 2007, and his MTech degree in engineering in 2011 from the Department of Electronics and Telecommunication Engineering at Veer Surendra Sai University of Technology, Burla, India. He received his PhD degree in electrical engineering from the National Institute of Technology, Rourkela, India in 2016. From 2016 to 2018, he worked for CV Raman College of Engineering, Bhubaneswar, India. Since 2018 he has been with the School of Electronics Engineering, KIIT University, Bhubaneswar, India as an assistant professor. His main research interests are wireless communication and body area networks.