IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 20, 2019, accepted January 24, 2020, date of publication January 30, 2020, date of current version February 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2970581

Interference Mitigation for Coexisting Wireless Body Area Networks: Distributed Learning Solutions

EMY MARIAM GEORGE, (Member, IEEE), AND LILLYKUTTY JACOB, (Senior Member, IEEE)
National Institute of Technology Calicut, Kozhikode 673601, India

Corresponding author: Emy Mariam George (emy_p160047ec@nitc.ac.in)

ABSTRACT When multiple wireless body area networks (WBANs) exist in close proximity to each other, the inter-user interference considerably degrades the signal to interference plus noise ratio of the packets arriving at each WBAN coordinator. Also, the propagation paths within each WBAN experience fading due to the continuous changes in the body posture and mobility of the human body. The most preferred coexisting mechanisms specified in the IEEE 802.15.6 standard is the channel hopping mechanism, which fails to consider the varying radio environment and obtained reward in its channel selection. Thus, our paper investigates this channel selection problem for interference mitigation in a time-varying environment. We formulate this channel selection problem as a finite repeated potential game and propose two learning algorithms, Stochastic Learning Algorithm (SLA) and Stochastic Estimator Learning Algorithm (SELA) to achieve the Nash Equilibrium (NE) of the game. Numerical results show the convergence of the learning algorithms to the NE point of the game. The performance evaluation and impact of parameters on these two algorithms are also analyzed in our paper.

INDEX TERMS Channel hopping, IEEE 802.15.6, interference mitigation, stochastic estimator learning algorithm, stochastic learning algorithm, potential game, Q-learning algorithm, WBAN.

I. INTRODUCTION

One of the most successfully emerged proactive health care systems, now in the world, is the wireless body area networks (WBAN). These networks can provide continuous health monitoring and real-time feedback to the user or medical personnel without hindering the user's lifestyle. A single WBAN consists of a number of sensor nodes and a coordinator node placed on different parts of a human body, which are capable of establishing wireless communication. The coordinator node is responsible for collecting, processing and transmitting the sensed physiological information. The perseverant interests in these networks have led to the development of the communication standard for WBAN, IEEE 802.15.6, in 2012.

IEEE 802.15.6 standard for WBAN defines the physical and medium access control (MAC) sublayers, which allow the devices (sensor nodes) to operate on low power, with low complexity, in or around the human body. The standard requires up to 10 coexisting WBANs to function properly in a 6 m^3 volume [1]. However, IEEE 802.15.6 does not specify any wireless channel access coordination among the WBANs at the MAC sublayer. This could lead to severe co-channel interference among the coexisting WBANs (see Fig.1). Also, the propagation paths within each WBAN experience fading due to different reasons such as energy absorption, reflection, diffraction, shadowing by body parts, changes in the body posture and mobility of the human body. So in a broader sense, our aim is to design and develop an efficient coexisting mechanism for multiple WBANs in a time-varying radio environment.

Most of the existing proposals for interference mitigation in other networks fail to become possible solutions, due to the dense deployment of users (WBANs), frequent topology changes in the network, and the mobility of the WBANs [2], [3]. The popular power control games adopted in cellular networks become less effective, due to low power consumption and group-based node structure of WBAN. Also, schemes developed for static and low mobility scenarios in a wireless sensor network (WSN) have been found unsuitable for WBANs. Thus, interference mitigation for coexisting

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

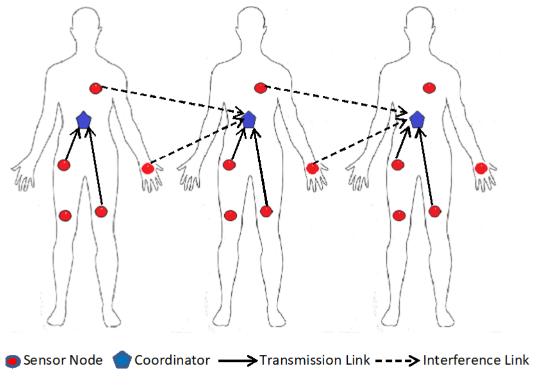

FIGURE 1. **Example of coexisting WBANs.**

WBANs becomes more challenging when compared to the cellular and wireless sensor networks [4].

The optional mechanisms specified in the standard for interference mitigation among coexisting WBANs are beacon shifting, channel hopping and active superframe interleaving [1]. Their applicability to the different operating frequency bands is also given in the standard. Coexisting WBANs can adopt one of the mechanisms based on its effectiveness, feasibility, and traffic volume in the network. In beacon shifting, WBANs can schedule the beacon transmission with backoff mechanism for an idle period. However, this method becomes ineffective if the number of coexisting WBANs is large because the performance would decline due to lack of idle period.

In channel hopping, the hub (coordinator) may change its operating frequency periodically by hopping to a randomly selected new channel after dwelling in the current channel for a fixed number of superframes as communicated to the sensor nodes. The operating frequency bands along with the numbers of channels available to each WBAN are given in the standard [1]. No message exchange among the coexisting WBANs is required in channel hopping and beacon shifting mechanisms, while superframe interleaving involves message exchange among the coexisting WBANs, resulting in null interference. From the standard, we know superframe interleaving is suitable only for the static environment and can cause considerable packet delay when the coexisting WBANs are more. Even the channel hopping mechanism may not be suitable for a dynamic environment if the number of channels in the frequency band is fewer than the coexisting WBANs.

Thus, the major challenges for interference mitigation of a dense system of WBANs are the following: (i) they are distributed; (ii) they exchange no information with their neighbors; (iii) they undergo block fading, i.e., the interference channel gain may remain constant for a slot but varies from slot to slot. Here, slot refers to the time duration for which a WBAN dwells in a channel before changing it. Our objective in this work is to design a mechanism that enables all the coexisting WBANs to choose those new channels for hopping that minimize the aggregate interference. Every WBAN selfishly tries to minimize its interference. This motivates us to solve the formulated optimization problem using game theory. But the existing game-theoretic solutions described in the 'Related Works' section work well only in a static radio environment and also they demand information from other players. So we need to consider distributed, uncoupled solutions that can adapt to time-varying changes in the environment and the coordinators can learn desirable information from their actions.

The main contributions of the paper can be summaried as follows:

- We formulate the channel selection of the coexisting WBAN coordinators for interference mitigation in a time-varying enviornment as an exact potential game, where utility is the weighted aggregate interference. It is shown that the channel selection (action) profile which globally minimizes the interference is a pure strategy Nash Equilibrium (NE) of the game.
- We consider two learning algorithms that can achieve the NE points of the formulated game. Both the algorithms are distributed and do not require information exchange between the players. Also, we study the performance of these algorithms in both static and dynamic environments. The convergence behavior of these algorithms are also analyzed.

The rest of the paper is organized as follows. Section II gives a brief description of related works. Section III explains the system model and defines the system utility. Section IV explains the channel selection as a potential game and how stochastic learning algorithm and stochastic estimator learning algorithm help in achieving NE in a time-varying environment. Section V presents the simulation results and the paper is concluded in Section VI.

II. RELATED WORKS

Several works have been done in the past years to address the severity of mutual interference in WBANs. Based on the technique adopted by the existing solutions, we can classify them into time spacing, frequency spacing, standard modifications, and hybrid solutions. Most of the works from [4]–[8] are based on time spacing where the simultaneous transmissions that interfere within the coexisting WBANs are avoided. Major drawbacks of these schemes are that all WBANs should be using same communication protocol, large number of WBANs could lead to large packet transmission delay, and scheduling of the transmission involves periodic exchange of information between the coexisting WBANs.

In [9], inorder to mitigate inter-WBAN interference the authors select a low power operation having suitable modulation scheme, data rate, and the duty cycle based on the measured SINR value. The hybrid solutions for interference mitigation are discussed in [10], [11]. In [10], first an interference prediction module estimates the interference based on distance and RSSI; and later a resource arbitrator allocates orthogonal time slots, frequencies, and codes to WBANs. A distributed mutual interference mitigation for Zigbee-based WBAN was proposed in [11] where each WBAN monitors the activity of all interfering WBANs in its vicinity. Based on the collected information, each WBAN reschedules its transmissions to empty time slots or switch to another idle channel.

Frequency spacing solutions in [12]–[14] use the frequency channels that are available for WBAN in [1]. These solutions allocate orthogonal channels either to an interfering sensor node or to interfering WBAN as a whole. The authors of [12] present a distributed interference mitigation mechanism for cluster tree Zigbee networks. On detecting high data and beacon collision, the nodes transmit the data on multiple channels determined by the coordinator. In [13] the channel assignments are broadcasted to allow all the WBAN devices to know the assigned channels for all their neighbors. In [14], authors propose channel switching based on the measured interference level. Even though frequency spacing schemes are highly suitable for dynamic environments, they face spectral inefficiency problem, and also are limited by the number of channels free of interference from networks like ZigBee, WiFi, etc.

Game theoretical solutions have been applied for interference mitigation in WBANs, which include power control game and least interference channel selection. The power control games consider both energy efficiency and coexisting interference of WBANs in their scheme. In [15] and [16], power control games make predictions of channel state according to the received information and adjust transmit power according to the predictions. Our own work in [17] proposes distributed learning for interference mitigation for WBANs in a static environment, where we formulate the channel selection problem as a finite repeated potential game and propose a distributed stateless Q-learning algorithm to achieve the Nash Equilibrium (NE). There we also propose an interference aware channel hopping mechanism for a coordinator to select the channel with the least interference as the most probable for transmission. The legacy frequency hopping involves three steps: channel measurements, channel classification, and channel hopping. A channel is classified into ‘good’ or ‘bad’ according to the predefined threshold and the good channels are used with a uniform hop probability. The selection of the predefined threshold does not take the number of interferers into account, thus making the legacy schemes less effective. The interference aware frequency hopping [17] dynamically classifies channels into ‘good’ or ‘bad’ according to the observed (real) interference levels and uses good channels with non-uniform hop probabilities. As an extension of our work in [17], we propose learning algorithm based solutions in a dynamic environment, in the present work.

III. SYSTEM MODEL

The system comprises a set of coexisting WBANs $\mathbb{N} = \{1, 2, \ldots, \mathrm{N}\}$ competing for a set of wireless channels $\mathbb{M} = \{1, 2, \ldots, M\}$ in a time-varying dense environment, forming a wireless network of WBANs with each WBAN acting as an autonomous entity distributed in space. Each WBAN is composed of K sensor nodes with different priorities and one coordinator (hub) placed in or around the human body, forming a one-hop star topology. In each WBAN, the hub shall operate in beacon mode with beacon periods, where the scheduled access and random access are carried out by TDMA and CSMA/CA, respectively. We assume that there is null intra-WBAN interference in the network. The coordinator node is responsible for selecting the channel used for intra-WBAN communication. The proposed system model is similar to the model considered in [18].

Practically, the co-channel interference between the coexisting WBANs decreases with an increase in their distances. To picture the limited range of interference, the system is represented by an undirected graph $\mathbb{G} = (\mathbb{V}, \mathbb{E})$ where $\mathbb{V}$ is the vertex set and $\mathbb{E}$ is the edge set. Each vertex corresponds to a WBAN. Two WBANs m and n are connected by an edge means that they interfere with each other when transmitting in the same channel; and the interference gain between them is denoted by w^{s}_{mn}, where s is the selected channel.

We assume that the available set of channels of each WBAN undergo block fading, i.e., the interference gains are constant in a time slot and vary randomly in the next time slot. Within a WBAN, the propagation paths can experience fading due to energy absorption, reflection, diffraction, shadowing by the environment surrounding the human body or by changes in the body posture. Other reasons for fading are the multipath propagation due to the environment around the body. In our model, we are considering the fading due to changes in the body posture in a hospital environment. The channel model for calculating the interference gains for different scenarios are described in [19] and are mentioned section V.

Let the set $\mathbb{G}_n$ represents the neighbors of WBAN n which is given by

$$\mathbb{G}_n = \{m \in \mathbb{N} : (m, n) \in \mathbb{E}\} . \quad (1)$$

We assume that each WBAN chooses exactly one channel for intra-WBAN communication at a time. Let $a_n \in \mathbb{M}$ be the channel selection of WBAN n. So, the channel selection profile of the N coexisting WBANs being $\{a_1, \ldots., a_N\}$, the interference experienced by WBAN n can be defined as

$$I_n = \sum_{m \in \mathbb{G}_n} p_m w^{a_n}_{mn} \delta(a_m, a_n) \quad (2)$$

where p_m is the average transmit power of the sensors of WBAN m; $\mathbb{G}_n$ as given by (1); and $\delta(a_m, a_n)$ is the Kronecker delta function given by

$$\delta(a_m, a_n) = \begin{Bmatrix} 1, & a_m = a_n \\ 0, & a_m \neq a_n \end{Bmatrix} . \quad (3)$$

With an aim to minimize the interference in the system, the system utility is defined as the weighted aggregate interference given by

$$U = \sum_{n \in \mathbb{N}} p_n I_n = \sum_{n \in \mathbb{N}} \sum_{m \in \mathbb{G}_n} p_n p_m w^{a_n}_{mn} \delta(a_m, a_n) . \quad (4)$$

Thus, our goal becomes minimizing the system utility given in (4). Choice of the weighted aggregate interference as the system utility results in balancing the transmit power and the interference experienced. It also leads to near-optimal system sum rate in the low SINR regime [18]. Its minimization can be achieved through optimal channel selection profile. The sum rate of the system is given by

$$r^{sum} = \sum_{n\in\mathbb{N}} r_n \tag{5}$$

where r_n is the rate of WBAN n in bps/Hz, which is calculated as

$$r_n = log_2\left(1 + \frac{1}{K}\sum_{k=1}^{K}\frac{g_n^k}{N_0 + I_n}\right) \tag{6}$$

where g_n^k is the received power at the hub of WBAN n from sensor k, N_0 is the noise power, and K is the number of sensor nodes in a WBAN. The received power at the hub of WBAN n from sensor k, can be calculated using Table 1. Thus, the optimization of the system utility given in (4) is an NP-hard problem even for a centralized system. Our aim in this work is to find an optimum dynamic channel selection profile $a^* = \{a_1^*, ..., a_N^*\}$ for minimizing the system utility given in (4).

IV. POTENTIAL GAME AND DISTRIBUTED LEARNING ALGORITHMS

Since the channel selection decision at each WBAN is carried out independently, each individual selection could affect the transmission of data in the neighbouring WBANs. This motivates us to formulate this as a game. So we consider a finite repeated potential game $\Psi = \left(\mathbb{N}, (A_n)_{n\in\mathbb{N}}, (u_n)_{n\in\mathbb{N}}\right)$, where $\mathbb{N}$ is the set of coexisting WBANs who are the players of the game and u_n is the utility function of WBAN n. The set of action profiles of all WBANs is given by $\mathbb{A} = \prod_{n\in\mathbb{N}} A_n$ where $A_n = \mathbb{M}$ is the set of available channels for player n and $\prod$ represents the Cartesian product. Generally the utility function of a WBAN is denoted as $u_n\left(a_n, \boldsymbol{a_{-n}}\right)$. However, the individual utility function of a WBAN is only dependent on its own channel selection and its neighbors. Thus the utility function can be reduced to $u_n\left(a_n, \boldsymbol{a_{\mathbb{G}_n}}\right)$, where $\boldsymbol{a_{\mathbb{G}_n}}$ is the channel selection of WBAN n's neighbours defined in (1). The individual utility function is chosen as

$$u_n\left(a_n, \boldsymbol{a_{\mathbb{G}_n}}\right) = p_n I_n \tag{7}$$

where p_n is the average transmit power of the sensors of WBAN n and I_n is given by (2). It should be noted that the experienced interference is a random variable in a slot. Thus, WBAN players experience random rewards in each slot.

Theorem 1: The formulated channel selection game is an exact potential game that has atleast one pure strategy NE. The optimal channel selection profile $\mathbf{a}^* \in \mathbb{A}$*, which globally minimizes the weighted aggregate interference is a pure strategy NE point of game* Ψ*.*

Proof: To prove this theorem, we consider the following potential function:

$$\phi = -\frac{1}{2}\sum_{n\varepsilon\mathbb{N}}\sum_{m\varepsilon G_n} p_n p_m w_{mn}^{a_n}\delta(a_m, a_n) \tag{8}$$

which can be rewritten as

$$\phi = -\frac{1}{2}U \tag{9}$$

where U is the system utility specified in (4). The change in the utility function of a player by unilaterally changing its channel selection is same as the change in the potential function. Thus the proposed channel selection game is an exact potential game with ϕ as its potential function. Suppose by contradiction that a^* is not a pure strategy NE. Therefore WBAN n can improve by deviating to a new profile $\hat{a}$ such that $u_n(\hat{a}) < u_n(a^*)$ which implies $\phi(\hat{a}) > \phi(a^*)$. Thus, contradicting that a^* maximises ϕ, or minimizes the weighted aggregate interference. ■

Important properties of an exact potential game are the following [18], [20]:

Property 1. *Any global or local maximum of the potential function is also a pure strategy NE of the game.*

Property 2. *Every improvement path in a potential game is finite.* The finite improvement path property ensures that the behavior of players who play best response in each iteration of the repeated game converge to a NE in finite time. Since we assume the interference is symmetrical (channel reciprocity), the channel with least amount of interference at WBAN m is also the the channel that generates least amount of interference at WBAN n. Thus, this selfish best response behavior by each player helps them to converge to a NE.

Property 3. *A potential game converges under round-robin, random and asynchronous timing. It does not converge under synchronous timing.* Since the coexisting WBANs are distributed networks, the probability of all WBANs choosing their channels at the same time is zero, thus convergence of the game is assured.

According to Theorem 1, the NE of the channel selection game coincides with the optimum channel selection profile, $\mathbf{a}^*$. Thus, now we need to consider a distributed learning algorithm that achieves one of the NE point's in the potential game. Most of the existing learning algorithms depend on the information from the other players while updating their actions and also requires the environment to be static. But, obtaining information from other WBANs is not possible and the radio environment varies from time to time due to changes in the human body posture. Thus in our work, we propose schemes based on stochastic learning algorithm (SLA) and stochastic estimator learning algorithm (SELA), which help to achieve pure strategy NE from their individual rewards. Also, we give a brief description of the Q-learning algorithm based scheme which we proposed earlier for a stationary environment [17].

A. DISTRIBUTED WBAN CHANNEL SELECTION USING SLA

In SLA, each WBAN chooses a channel according to the mixed strategy probability distribution of the set of available channels, $\mathbb{M}$. Depending on its selected channel, each WBAN receives a random payoff. Then the WBAN updates its mixed strategy of channel selection based on the received payoff. If the obtained payoff is rewarding, then the probability of selection of this channel in the next iteration increases. The payoff obtained at the slot t on choosing a channel by WBAN n is

$$R_n(t) = \frac{r_n(t)}{r_n^*} \tag{10}$$

where $r_n(t)$ is the achieved rate of WBAN n at the time t and r_n^* is the maximum achievable rate when there is no interference. Such a payoff is used for distributed learning based resource allocation in wireless networks [21]. The achieved rate of WBAN n is given in (6) and $r_n^* = log_2\left(1 + \frac{1}{K}\sum_{k=1}^{K} \frac{g_n^k}{N_0}\right)$. When the variation in the system utility during a period is trivial, the learning is stopped.

The convergence of the SLA is given in [22] which states that with sufficiently small size for b, the learning algorithm asymptotically converges to pure strategy NE of the game.

B. DISTRIBUTED WBAN CHANNEL SELECTION USING SELA

With most of the classical learning algorithms, the players update the probability vectors directly based on their instantaneous payoffs, as we have seen with SLA. If the payoff of an action is rewarding, then the probability of selecting that action in the next iteraion is increased. Otherwise, the probability of that action remains unchanged or decreased. In a time varying environment, the probability of the validity of an action based on the 'old response' decreases. So we consider another learning algorithm, which utilizes a stochastic estimator to operate in the non-stationary environment [23]. This algorithm is characterized by the indirect use of mean payoffs of each action in the probability updation.

Algorithm 1 Stochastic Learning Algorithm (SLA)

```
Intialisations: Set $t = 0$ and set the initial mixed strategy of each WBAN to $p_{ns}(t) = 1/|A_n|, \forall n \epsilon \mathbb{N}, \forall s \epsilon \mathbb{M}$.
Loop for $t = 0, 1, 2, 3..$
1. In the $t$th slot, each WBAN selects a channel $a_n(t)$ according to its channel selection probability vector $p_n(t)$.
2. Based on the selected action, each WBAN calculates the received payoff using (10).
3. All the WBANs update their mixed strategy according to the following rule:
    $p_{ns}(t+1) = p_{ns}(t) + bR_n(t)(1 - p_{ns}(t)), \quad s = a_n(t)$
    $p_{ns}(t+1) = p_{ns}(t) - bR_n(t)p_{ns}(t), \quad s \neq a_n(t) \quad (11)$
where $0 < b < 1$ is the learning step size and $R_n(t)$ is the normalised received payoff as defined in (10).
```

The stochastic estimator estimates the mean rewards of each action and doubts the validity of the environmental response by adding a zero mean normal distributed random number to the mean reward of each action. The variance of this normal distribution is directly proportional to the time elpased from the instant when that action was last selected. Thus, it improves the probability of actions that have not been selected recently to be estimated as 'optimal'. So, in stochastic estimator even when an action is rewarded, it is possible that probability of choosing another action is increased; because this learning algorithm increases the probability of the action that has the highest estimated mean reward. The SELA is powerful, flexible and ergodic, i.e., converges at the optimal action with a distribution independent of the intial state; and is defined by $< A, B, P, T, E >$ [23] where:

- A is the set of M actions available to each WBAN user.
- B is the set of possible received payoffs by a WBAN, corresponding to the set of M possible actions. Since we consider a normalised payoff in (10), the payoff at any time belongs to [0, 1].
- P is the probability vector of choosing each action. $P_n(t) = \{p_{n1}(t), p_{n2}(t), \ldots, p_{nM}(t)\}$, where p_{ns} is the probability of choosing action s by WBAN n, i.e,. $a_n(t) = s$.
- T is the learning algorithm described in Algorithm 2 that modifies the probability vector $P_n(t)$ at each iteration using the calculated payoffs.
- E is the estimator which is defined by $E(t) = (D'(t), M(t), U(t))$. $D'_n(t)$ is the Deterministic Estimator Vector for WBAN n which contains the current deterministic estimates of the mean rewards of the actions. The mean reward of action s is defined as

$$d'_{ns}(t) = \frac{Q}{\mathrm{W}} \tag{12}$$

where Q is the total payoff received the last W times action s was selected and W is called the 'learning window'. $M_n(t)$ is the Oldness Vector for WBAN n which contains the time which has elapsed from the last time each action was selected. $m_{ns}(t) \in M_n(t)$ of action s is given as

$$m_{ns}(t) = t - \underset{t'}{max}\left\{t' : t' \leq t \quad and\ a_n\left(t'\right) = s\right\} \tag{13}$$

$U_n(t)$ is the Stochastic Estimator Vector of WBAN n which contains the current stochastic estimates of the mean payoffs of the actions. $u_{ns}(t) \in U_n(t)$ of action s is given as

$$u_{ns}(t) = d'_{ns}(t) + \mathcal{N}\left(0, \sigma_s^2(t)\right) \tag{14}$$

where $\sigma_s(t) = min\{\alpha m_{ns}(t), \sigma_{max}\}$. $\mathcal{N}\left(0, \sigma_s^2(t)\right)$ is a random number from normal distribution with 0 mean and variance equal to $\sigma_s^2(t)$. α is a parameter that determines how fast the stochastic estimates deviate from deterministic estimates, and σ_{max}^2 bounds the variance of the stochastic estimates preventing it from increasing infinitely.

Theorem 2: The Stochastic Estimator Learning Algorithm is ε-optimal in every stochastic environment that offers symmetrically distributed noise. Thus, if action m is the optimal one and $p_{nm}(t) = Pr[a_n(t) = m]$, then for every value $R \geq R_0 (R_0 > 0)$ of the resolution parameter there is a time instant $t_0 < \infty$ such that for every $t \geq t_0$ it holds that $E\left[p_{nm}(t)\right] = 1$

Proof: Proof of this theorem is given in [23]. ■

C. Q-LEARNING ALGORITHM

As mentioned earlier, we also consider a standard reinforcement technique, *viz.* Q-learning, that enables the WBANs to learn their optimal channel selections in a static environment.

Like in SLA and SELA, the Q-learning algorithm doesn't assume any knowledge for the player about its environment, rather player must learn from its environment. On performance of each action, a reinforcement signal is generated that is then used to evaluate the performed actions by updating its Q-value. Each player of the game applies the Q-learning independently by ignoring the action selection by the other players.

Specifically, after playing the action a_n at iteration t and recieving a reward of $R_n(t)$, WBAN n updates its Q-table using the equation

$$Q_n^{t+1}(a_n) = Q_n^t(a_n) + \lambda(t)(R_n(t) - Q_n^t(a_n)) \tag{16}$$

where $\lambda(t) \in [0, 1]$ is a learning parameter. The convergence of Q-learning with probability 1 happens if the conditions $\sum_{t=1}^{\infty} \lambda(t) = \infty$ and $\sum_{t=1}^{\infty} (\lambda(t))^2 = \infty$ hold. i.e., if the Q-values are updated infinitely often. We use the following form of $\{\lambda\,(t)\}_{t\geq 1}$ in our work:

$$\lambda(t) = (\beta + \Delta^t(a_n))^{-\rho} \tag{17}$$

where β is an arbitrary positive constant, $\Delta^t(a_n)$ is the number of times the action a_n has been selected upto time t and $\rho \in [0, 1]$ is the learning rate parameter. The condition that all actions are performed infinitely can be met using a randomized policy, in which the probability of playing each action is bounded by a sequence that tends to zero sufficiently slowly as t becomes larger. Here we consider ϵ-greedy action selection strategy, where the probability of selecting maximal reward actions tends to 1 as t tends to ∞. The ϵ value is updated as [24]:

$$\epsilon_t = \epsilon_0 t^{-1/N} \tag{18}$$

A player either selects a greedy action at time t with probability of $(1 - \epsilon_t)$ or chooses to explore by selecting an action randomly from $\mathbb{M}$ with probability ϵ_t. So the action selection in a WBAN can be given by

$$a_n^t = \left\{ \begin{array}{cc} arg\,max\, Q_m^t(a_m), & w.p.\, 1 - \epsilon_t \\ Uniform(1, 2, \ldots, M), & w.p.\, \epsilon_t \end{array} \right\} \tag{19}$$

Theorem 3: For the potential game Ψ, there exists an $\epsilon > 0$ such that under Q-learning with an action selection strategy as given in (19), for sufficiently large t*, the probability that a_n^t is a NE is 1.*

Proof: Given in [24]. ■

So in our game model of coexisting WBANs, WBAN n selects a channel $a_n \in \mathbb{M}$ independently and receives an individual reward. The payoff obtained at the iteration t on choosing a channel $a_n \in \mathbb{M}$ is

$$R_n(t) = \frac{r_n(t)}{r_n^*} \tag{20}$$

as defined in (10) and can be calculated using local measurements. The execution at each WBAN is as follows:

- WBAN n sets $Q_n^t\,(a_n) = 0, \forall a_n \epsilon \mathbb{M}$.
- At each iteration:
  - selects its action by following ϵ-greedy algorithm given in (19). i.e. It either selects the channel with the best Q-value with a probability of $1 - \epsilon_t$ or selects a channel uniformly from the action set $\mathbb{M}$ with a probability of ϵ_t.
  - computes the reward $R_n(t)$ and updates $Q_n^{t+1}\,(a_n)$ according to (16) using the learning parameter in (17).
  - ϵ_t is updated as per (18).

The iteration is continued until the convergence is obtained.

Algorithm 2 Stochastic Estimator Learning Algorithm (SELA)

```
Intialisations: Set $t = 0$, and set the initial probability vector of each WBAN to $p_{ns}\,(t) = 1/\,|A_n|$, and $d'_{ns}(t) = m_{ns}(t) = u_{ns}(t) = 0, \forall n \epsilon \mathbb{N}, \forall s \epsilon \mathbb{M}$.
Loop for $t = 0, 1, 2, 3..$
1. Select the action $a_n(t) = s$ according to the probability vector.
2. Calculate the payoff received for the selected action $s$ according to (10).
3. Compute the new deterministic estimate for the action $s$ by (12).
4. Update the oldness vector of the selected action $s$ by setting $m_{ns}(t) = 0$ and for other remaining actions, oldness vector is updated according to (13); i.e., $m_{ni}(t) = m_{ni}(t-1)+1, \forall i \neq s$.
5. For every action, compute the new stochastic estimate $u_{ns}(t)$ according to (14). Identify the optimal action $'m'$ having the highest stochastic estimate of mean reward.
6.Update the probability vector in the following way:
```

$$p_{ns}(t+1) = \left\{ \begin{array}{cc} p_{ns}(t) - 1/R, & s \neq m \\ 1 - \sum_{s \neq m} p_{ns}(t+1) & s = m \end{array} \right\} \tag{15}$$

where R is called the resolution parameter.

TABLE 1. Path loss model for body surface to body surface CM3 channel at 2.4GHz [19].

Path Loss	Hospital Room
	$PL(dB) = alog_{10}(d) + b + N$
a	6.6
b	36.1
σ_N	3.80

a and b are coefficients of linear fitting, d is the transmitter-receiver distance in mm and N is a normally distributed variable with zero mean and standard deviation σ_N

TABLE 2. Scenarios and the distributions that give the best fit to those scenarios transmitting from right wrist to receiver off the body at 2.36 GHz [19].

Action	D (in m)	Angle	Distribution
Stand	1	0	Weibull (a=0.97,b=60.7)
Stand	2	0	Lognormal (μ=-0.051,σ=0.018)
Stand	3	0	Lognormal(μ=-0.13,σ=0.031)
Stand	4	0	Lognormal(μ=-0.77,σ=0.33)

Angle is orientation of the subject with respect to the receiver in degrees, D is horizontal distance from subject to receiver

V. SIMULATION RESULTS

We consider a dense multi-WBAN system of 50 WBANs randomly located in a $10*10m^2$ hospital region. Even though we assume that the WBAN's are operating in the 2.4GHz band, we consider availability of only 10 channels of 1MHz bandwidth for channel selection. The reason for this assumption is that 2.4GHz band is shared among many technologies like Bluetooth, Bluetooth Low-Energy, ZigBee, WiFi, etc. But in our problem formulation for interference mitigation, we are only considering interference from coexisting WBANs. Thus we are restricting our available channels to the number of non-over lapping channels of WBAN's 2.4GHz band and WiFi's channels 1, 6, 11. The transmission power of all WBAN communications is set as $p_n = 0$ dBm, $\forall n \epsilon \mathbb{M}$ and the noise power as $N_0 = -70$dBm. Also in each WBAN, we assume the sensor node locations as random on the human body at distance from the coordinator ranging from 100 to 1000 mm.

The channels are assumed to undergo block fading, i.e., the fading remains constant in an iteration and randomly changes in the next iteration. The intra-WBAN received power and inter-WBAN interference at each coordinator are calculated using respectively, CM3 and CM4 channel models in [19]. The parameters for path loss model for body surface to body surface CM3 channel at 2.4GHz is given in Table 1. In [19], they have also compiled a list of 'best distribution fit' for the normalized received power for on-body to off-body communications with all scenarios like the subject walking or standing still, with varying distance and angle of orientation. It is seen that the Lognormal distribution is the best matching model of normalized received power while the subject is standing still, independent of the location of the antenna on the human body. Table 2 shows a few scenarios and their distribuions that we have considered in our work.

A. CONVERGENCE BEHAVIOUR

For studying the convergence behavior of the considered algorithms, we focus on a random WBAN from the dense multi-WBAN system. All the 10 channels undergo fading as described in Tables 1 and 2. The convergence behavior of SLA and SELA for a random trial is shown in Fig.2 and Fig.3, respectively. In both the learning algorithms, the initial channel selection probabilities of all the channels are 0.1. From Fig. 2, it can be seen that SLA converges at around 250 iterations. We can see that the random user starts selecting channel 5 with probability 1. Similar conclusions can be drawn for SELA from Fig. 3. It can be seen that the user starts selecting channel 3 with probability 1 from 200 iterations onwards.

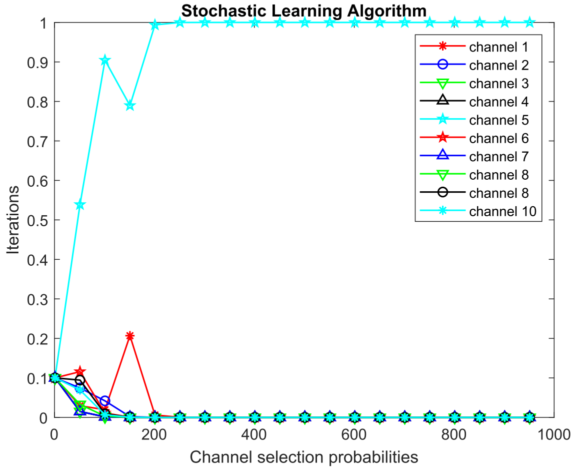

FIGURE 2. Evolution of channel selection probabilities in the SLA.

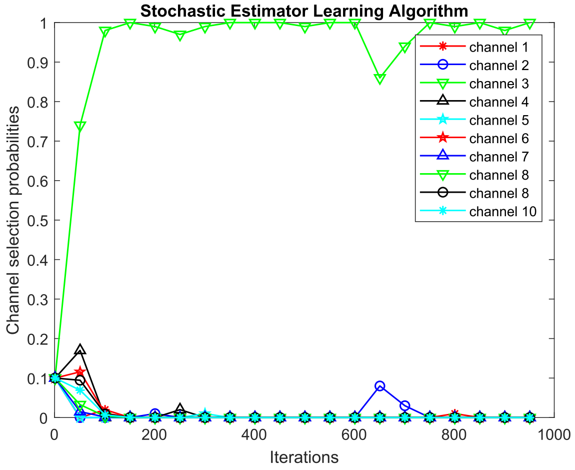

FIGURE 3. Evolution of channel selection probabilities in the SELA.

B. PERFORMANCE EVALUATION

We evaluate the performance of SLA and SELA in terms of weighted aggregate interference defined in (4). We consider a dense network of 50 WBANs with a fixed set of values for the LA parameters first. The learning step size b of SLA is set as 0.3. The parameters affecting the SELA algorithm are W, R, α, σ_{max} which are set as 10, 100, 0.001, 1, respectively. The random channel selection scheme of IEEE 802.15.6 standard is also considered for comparison purpose. In random channel selection, each WBAN user randomly selects a channel in

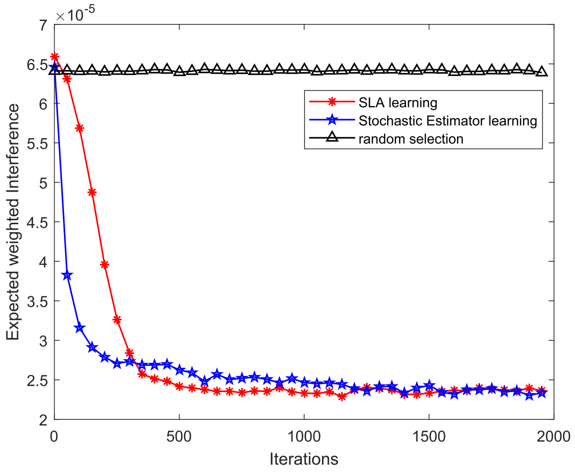

FIGURE 4. Convergence Behavior of Expected Weighted Interference vs. Iterations for SLA, SELA and random selection in time varying environment.

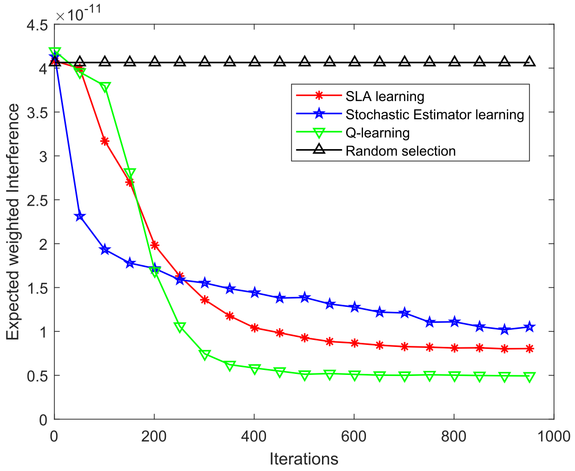

FIGURE 5. Convergence Behavior of Expected Weighted Interference vs. Iterations for SLA, SELA, Q-learning and random selection in static environment.

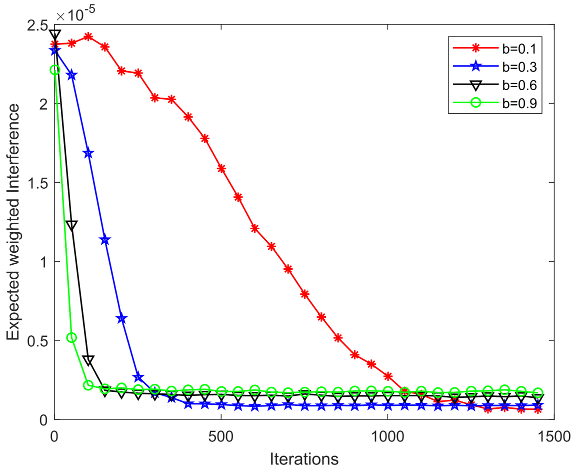

FIGURE 6. Impact of learning step size b in the SLA.

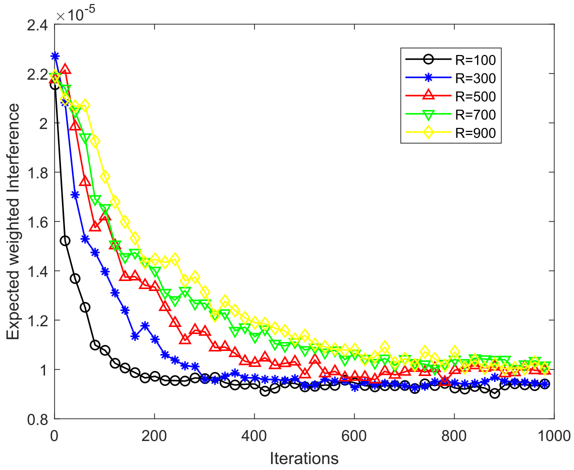

FIGURE 7. Impact of resolution parameter R in SELA.

each slot. From Fig. 4, it can be seen that the performance of the SELA algorithm matches up with SLA algorithm at higher iterations. Since the interference varies from slot to slot, the random channel selection gives the worst performance when compared with the learning algorithms which is seen in Fig. 4. Since convergence of Q-learning is not assured in dynamic environment, it is not considered for comparison.

Fig.5 shows the expected weighted iterference of the network in a static environment, by considering random channel hopping, SLA, SELA and Q-learning algorithm. The path loss measurements for the static environment is given in [19]. It can be seen from figure that the distributed Q-learning algorithm performs best when compared to the other three schemes in a static environment. It should be noted that all the three learning algorithms perform much better than the random hopping of IEEE 802.15.6 standard.

C. IMPACT OF PARAMETER VALUES

In this section, we study sensitivity of the convergence and performance of the learning algorithms to the parameter values. Fig. 6 studies the effect of learning step size b on SLA. It can be seen that selecting smaller b values greatly reduces the convergence speed of the algorithm and selecting higher b values greatly increase the convergence speed by trapping the algorithm to a local optimum value.

Fig. 7 and Fig. 8 show the impact of resolution parameter R and the internal parameter α, respectively in SELA algorithm. In Fig. 7, by selecting large resolution parameter R, it gives small probability updating factor which results in low convergence speed to the algorithm. From Fig. 8, it can be seen that selecting larger values of α result in the addition of a random number from a normal distribution with large variance to the mean reward. This leads to similar action selection behaviour as that of random channel hopping. An α value of 0.001 gives the least weighted interference with good convergence. Fig. 9 shows the impact of σ_{max} in SELA algorithm. From the figure, it is clear that changing the value of σ_{max} doesn't have much effect in the expected aggregate interference in our model.

Fig. 10 shows the effect of ϵ_0 in Q-learning algorithm on the system utility. The best results for the system utility is obtained when $\epsilon_0 = 0.55$, and the algorithm converges to the

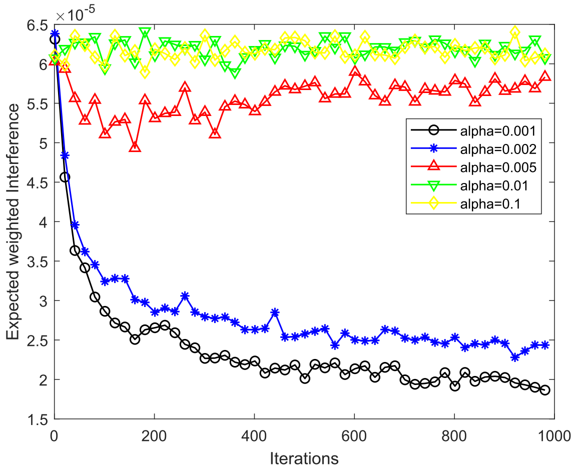

FIGURE 8. Impact of scaling parameter α in the SELA.

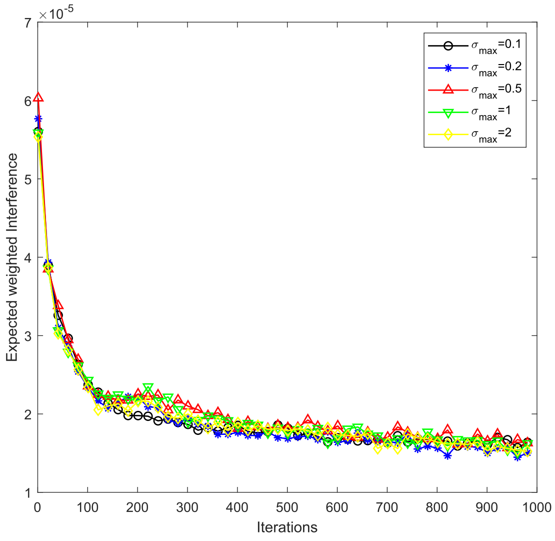

FIGURE 9. Impact of the σ_{max} in the SELA.

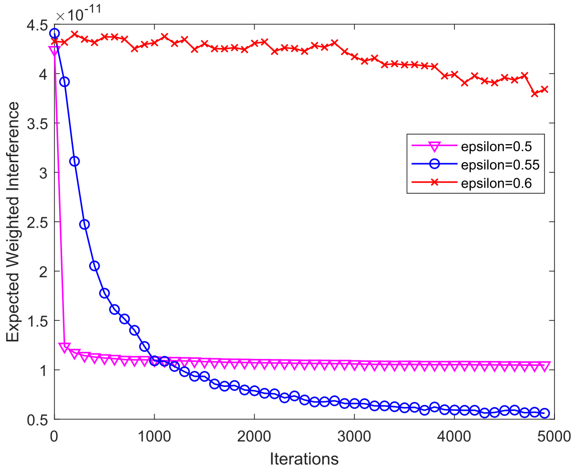

FIGURE 10. Effect of ϵ_0 on system utility in Q-learning algorithm.

optimal at around 2000 iterations. This means that by setting $\epsilon_0 = 0.55$ sufficient exploration of all channels are carried out before exploiting the maximal rewarding Q values. On setting $\epsilon_0 < 0.55$ the network settles into a higher system utility over time. This is due to insufficient exploration of all channels before exploitation. On setting $\epsilon_0 > 0.55$, the time required to converge to optimal value is very high.

VI. CONCLUSION

In this paper, we proposed the distributed learning based schemes for interference mitigation of coexising WBANs in a dynamic environment. We formulated the channel selection for hopping by individual WBANs as an exact potential game, where utility is the weighted aggregate interference. It is shown that the channel selection (action) profile, which globally minimizes the interference, is a pure strategy NE of the game. In order to achieve this NE, we considered two learning algorithms that are appropriate for dynamic radio environment, SLA and SELA. The near-optimal performance and convergence behavior of both the algorithms were evaluated and found to be better than the random channel selection specified in the IEEE 802.15.6 standard. The performance of these algorithms were also compared with Q-learning algorithm, which cannot be used for dynamic environment, for a static environment. One possible extension of this work is interference mitigation in a time-varying network topology where the number of active WBAN users varies.

REFERENCES

[1] *IEEE Standard for Local and Metropolitan Area Networks—Part 15.6: Wireless Body Area Networks*, Standard 802.15.6-2012, Feb. 2012, pp. 1–271.

[2] S. Movassaghi, M. Abolhasan, J. Lipman, D. Smith, and A. Jamalipour, "Wireless body area networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1658–1686, 3rd Quart., 2014.

[3] M. Chen, S. Gonzalez, A. Vasilakos, H. Cao, and V. Leung, "Body area networks: A survey," *Mobile Netw. Appl.*, vol. 16, pp. 171–193, Apr. 2011.

[4] S. Movassaghi, M. Abolhasan, D. Smith, and A. Jamalipour, "AIM: Adaptive Internetwork interference mitigation amongst co-existing wireless body area networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Austin, TX, USA, Dec. 2014, pp. 2460–2465.

[5] J. Dong and D. Smith, "Cooperative body-area-communications: Enhancing coexistence without coordination between networks," in *Proc. IEEE 23rd Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sydney, NSW, Australia, Sep. 2012, pp. 2269–2274.

[6] E.-J. Kim, S. Youm, T. Shon, and C.-H. Kang, "Asynchronous internetwork interference avoidance for wireless body area networks," *J. Supercomput.*, vol. 65, no. 2, pp. 562–579, Aug. 2013.

[7] S. Kim, S. Kim, J.-W. Kim, and D.-S. Eom, "A beacon interval shifting scheme for interference mitigation in body area networks," *Sensors*, vol. 12, no. 8, pp. 10930–10946, Aug. 2012.

[8] L. Wang, C. Goursaud, N. Nikaein, L. Cottatellucci, and J.-M. Gorce, "Cooperative scheduling for coexisting body area networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 123–133, Jan. 2013.

[9] W.-B. Yang and K. Sayrafian-Pour, "Interference mitigation using adaptive schemes in body area networks," *Int. J. Wireless Inf. Netw.*, vol. 19, no. 3, pp. 193–200, Sep. 2012.

[10] B. D. Silva, A. Natarajan, and M. Motani, "Inter-user interference in body sensor networks: preliminary investigation and an infrastructure-based solution," in *Proc. 6th Int. Workshop Wearable Implant. Body Sensor Netw.*, Berkeley, CA, USA, Jun. 2009, pp. 35–40.

[11] W. Sun, Y. Ge, and W.-C. Wong, "A lightweight inter-user interference mitigation method in Body Sensor Networks," in *Proc. IEEE 8th Int. Conf. Wireless Mobile Computing, Netw. Commun.(WiMob)*, Barcelona, Spain, Oct. 2012, pp. 34–40.

[12] J. Han, H. Kim, J. Bang, and Y. Lee, "Interference mitigation in IEEE 802.15.4 networks," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Kathmandu, India, Dec. 2011, pp. 1–5.

[13] W. Lee, S. H. Rhee, Y. Kim, and H. Lee, "An efficient multi-channel management protocol for Wireless Body Area Networks," in *Proc. Int. Conf. Inf. Netw.*, Chiang Mai, India, Jan. 2009, pp. 1–5.

[14] J. Mahapatro, S. Misra, M. Manjunatha, and N. Islam, "Interference-aware channel switching for use in WBAN with human-sensor interface," in *Proc. 4th Int. Conf. Intell. Human Comput. Interact. (IHCI)*, Kharagpur, India, Dec. 2012, pp. 1–5.

[15] W. Liu, J. Liu, and H. Zhao, "Power control mechanism based on hybrid access schemes in coexisting WBANs," in *Proc. 8th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Yangzhou, China, Oct. 2016, pp. 1–5.

[16] Y. Xu, M. Ke, F. Liu, and Q. Zha, "A self-adaptive Power control algorithm based on game theory for inter-WBAN interference mitigation," in *Proc. 2nd IEEE Int. Conf. Comput. Commun. (ICCC)*, Chengdu, China, Oct. 2016, pp. 2873–2877.

[17] E. M. George and L. Jacob, "Distributed learning algorithm for interference avoidance in coexisting WBANs," in *Proc. 2018 Int. Conf. Signal Process. Commun. (SPCOM)*, Bengaluru, India, Jul. 2018, pp. 477–481.

[18] Q. Wu, Y. Xu, J. Wang, L. Shen, J. Zheng, and A. Anpalagan, "Distributed channel selection in time-varying radio environment: Interference mitigation game with uncoupled stochastic learning," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4524–4538, Nov. 2013.

[19] K. Y. Yazdandoost and K. Sayrafian-Pour, *Channel Model for Body Area Network (BAN)*, Standard IEEE P802.15-08-0780-09-0006, Apr. 2009.

[20] J. Wang, Y. Xu, A. Anpalagan, Q. Wu, and Z. Gao, "Optimal distributed interference avoidance: Potential game and learning," *Trans. Emerg. Tel. Tech.*, vol. 23, no. 4, pp. 317–326, Jun. 2012.

[21] F. Wilhelmi, B. Bellalta, C. Cano, and A. Jonsson, "Implications of decentralized Q-learning resource allocation in wireless networks," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Montreal, QC, Canada, Oct. 2017, pp. 1–5.

[22] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y.-D. Yao, "Opportunistic spectrum access in unknown dynamic environment: A game-theoretic stochastic learning solution," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1380–1391, Apr. 2012.

[23] A. V. Vasilakos and G. I. Papadimitriou, "A new approach to the design of reinforcement schemes for learning automata: Stochastic estimator learning algorithm," *Neurocomputing*, vol. 7, no. 3, pp. 275–297, Apr. 1995.

[24] A. C. Chapman, D. S. Leslie, A. Rogers, and N. R. Jennings, "Convergent learning algorithms for unknown reward games," *SIAM J. Control Optim.*, vol. 51, no. 4, pp. 3154–3180, Jan. 2013.

EMY MARIAM GEORGE (Member, IEEE) received the B.Tech. degree in electronics and communication engineering from the Rajiv Gandhi Institute of Technology, Kottayam, India, in 2012, and the M.Tech. degree in communication engineering from the Vellore Institute of Technology, Vellore, India, in 2015. She is currently pursuing the Ph.D. degree with the Department of Electronics and Communication Engineering, National Institute of Technology Calicut, Calicut, India. Her current research interests include resource allocation, game theory, and wireless body area networks.

LILLYKUTTY JACOB (Senior Member, IEEE) received the Ph.D. degree in electrical communication engineering from the Indian Institute of Science Bangalore, in 1993. She was a Postdoctoral Fellow with the Korea Advanced Institute of Science and Technology, from 1996 to 1997. She was a Visiting Faculty with the National University of Singapore, from 1998 to 2003, and with the University of South Florida, in 2012. She is currently a Professor with the Department of Electronics and Communication Engineering, National Institute of Technology Calicut. She has received more than 160 publications in international journals and conferences. She is an editor and a reviewer for several international journals and conferences. Her research interests are in protocol design, performance modeling, and analysis of communication networks. She is a Life Member of ISTE and a Fellow of IEI and IETE.