A JOINT DUTY CYCLE SCHEDULING AND ENERGY AWARE ROUTING APPROACH BASED ON EVOLUTIONARY GAME FOR WIRELESS SENSOR NETWORKS

M.S. KORDAFSHARI, A. MOVAGHAR AND M.R. MEYBODI

Abstract. Network throughput and energy conservation are two conflicting important performance metrics for wireless sensor networks. Since these two objectives are in conflict with each other, it is difficult to achieve them simultaneously. In this paper, a joint duty cycle scheduling and energy aware routing approach is proposed based on evolutionary game theory which is called DREG. Making a trade-off between energy conservation and network throughput, the proposed approach prolongs the network lifetime. The paper is divided into the following sections: Initially, the discussion is presented on how the sensor nodes can be scheduled to sleep or wake up in order to reduce energy consumption in idle listening. The sensor wakeup/sleep scheduling problem with multiple objectives is formulated as an evolutionary game theory. Then, the evolutionary game theory is applied to find an optimal wakeup/sleep scheduling policy, based on a trade-off between network throughput and energy efficiency for each sensor. The evolutionary equilibrium is proposed as a solution for this game. In addition, a routing approach is adopted to propose an energy aware fuzzy logic in order to prolong the network lifetime. The results show that the proposed routing approach balances energy consumption among the sensor nodes in the network, avoiding rapid energy depletion of sensors that have less energy. The proposed simulation study shows the more efficient performance of the proposed system than other methods in term of network lifetime and throughput.

1. Introduction

A wireless sensor network (WSN) consists of a large number of tiny sensors [2]. Each sensor node is equipped with a radio transceiver, processing unit and power-constrained battery as an energy source. The WSNs are used in a wide range of applications [5] such as military and civil applications including target tracking, weather, building, environment monitoring, traffic control, health applications etc. It is very difficult to replace or recharge batteries of sensors in WSNs. Therefore, the sensors have limited energy, resulting in the limited lifetime of WSN. The lifetime of a WSN is most commonly defined as the time until a first sensor dies [11]. So energy conservation is one of the most important issues in WSNs. In the recent years, many methods have been proposed to conserve sensors energy (e.g.

Received: April 2016; Revised: July 2016; Accepted: October 2016

Key words and phrases: Wireless Sensor Network, duty cycle scheduling, Energy aware routing, Evolutionary Game Theory, Distributed Reinforcement Learning.

Energy aware routing [22], data aggregation [25], duty cycle scheduling [4]). In energy aware routing approaches, algorithms choose a suitable forwarding path, based on residual energy of path. Data aggregation methods reduce the number of transmitting packet, leading to less energy consumption. In duty cycle schemes, sensors enter sleep mode, where energy consumption is reduced since there is no traffic. A review of energy conservation solutions for WSNs was proposed in [26]. However, the main challenge is the conflict among the objectives. Therefore, there needs to be a trade-off between energy conservation and Quality of Service. This makes the problem of energy conservation more complicated [3].The majority of proposed techniques in literature are designed, based on optimization methods. These optimization approaches usually entail the presentation of repeated solution of the differential equation models. Therefore, they may be inefficient in practice. In addition, optimization algorithms can be used for modeling multi-agent systems, where agents as decision makers may set conflicting objectives, competing against each other. Each decision maker seeks its own interests in a self-appraisal scenario. Thus, efficient and flexible algorithms should be developed to model the dynamic and competitive behaviors of agents so that the total satisfaction level of the agent is maximized.

Game theory [35] can be applied as a useful tool to analyze complex decision making problems, where decision makers have the conflicting objectives. Many game-theoretic models [19] have been applied to the analysis of different problems of WSN such as routing [28],[14], duty cycle scheduling [34], quality of service [6], topology control [27] and etc.

The majority of the proposed methods in this area have been proposed as non-cooperative game, which are very helpful in WSN problem solving and are based on rationality assumptions. The players are rational in the sense that they always seek to maximize their own utilities [10]. Thus, each player can predict the action of others and choose an optimal strategy. The rationality assumption is often unrealistic. It is difficult to imagine fully rational decision making because of insufficient information (even in simple scenarios).The rationality assumption is relaxed in evolutionary game theory (EGT) [30]. The EGT studies the evolution of the population toward the optimal strategies through time. Evolutionary game involves a large number of players that form a population. In evolutionary game, the rationality of players is bounded. Players adopt their strategies by a trial-and-error method through repeating game until they learn their own best strategies over time. In each period of game, player chooses a strategy and calculates the resulting outcome. Then, the utility of population is observed. The player changes its strategy if his utility is less than the average utility of population.

In this paper, for the first time, we study an evolutionary game for duty cycle scheduling and energy conservation problem. We model the problem as a multi agent non-cooperative game. The provided evolutionary game theory technique captures the slow change of sensor duty cycle over time. As a result, each sensor repeatedly chooses a strategy that maximizes its utility. The game is repeated until a steady state, i.e., an evolutionary equilibrium, is reached. Unlike a classical non-cooperative game, in which all of the players make decisions which immediately lead to the sub optimal solution, the new proposed evolutionary game involves a situation where players slowly change their strategies to achieve the solution eventually.

The main contributions of this paper are summarized as follows:

- A joint duty cycle scheduling and energy aware routing scheme are proposed for network lifetime maximization. The proposed approach makes a suitable trade-off between the energy conservation and the throughput in each sensor. A selfish sensor decides whether or not to wake up so that the cycle obtains more utility in term of energy conservation and throughput.
- We formulate the wakeup/sleep scheduling as an evolutionary game theory.The EGT studies the evolution of the population of sensors toward the optimal strategies through time. This game yields the evolutionary equilibrium as a solution to achieve the optimal wakeup/sleep scheduling.
- Then, an energy aware routing algorithm is proposed. This algorithm in each sensor makes routing-related decision independently, based on local information. A sensor selects an active node with the highest residual energy among the group as a next hop. Considering the residual energy of sensors in routing decisions, this method balances the energy consumption in the network.
- Q-learning algorithm is used to implement the proposed evolutionary game-based model for duty cycle scheduling.

The remainder of this paper is organized as follows: Section 2 reviews related works. The system model and assumptions for the energy consumption and routing are discussed in Section 3. Section 4 presents an overview of evolutionary game theory, followed by the formulation of an evolutionary game model for the sensor wakeup/sleep scheduling problem. The Q-learning algorithm for duty cycle scheduling is presented in Section 5. Section 6 evaluates the performance of the proposed approach. Finally, section 7 Concludes our work.

2. **Related works**

Energy conservation to maximize the network lifetime is the main challenge in WSNs. Many routing protocols have been designed to reduce energy consumption and prolong the network lifetime. In [9], authors consider the residual energy of a sensor as a metric in routing decisions. They define three neighbor sensor types: parent sensors, sibling sensors and child sensors. Each sensor selects its parent to forward packet if its parent has the highest residual energy among the nearest neighbor sensors. If there is no parent sensor available, the sensor forwards the packet to the sibling sensor with the highest residual energy. This algorithm balances the energy consumption through the network to increase the network lifetime. Author in [24] defines sensors in the network as bottleneck nodes. The bottleneck nodes are sensors which consume their energy faster and die earlier than other sensors in the network. In this approach, the algorithm selects a sensor as a forwarding node. If the selected sensor is a bottleneck node, the sensor doesnt forward the packet to it but sends packet directly to the sink.

Authors in [15] proposed a game theoretic distributed Energy Balanced Routing (GTEBR) approach that prolongs the network lifetime. In this game, each sensor has a mixed strategy space as $S = \{0, 1\}$. The value of 1 means that sensor makes decision to be part of the route and forwards the packet. The value of 0 means that sensor decides not to play any role in forwarding packet and remains silent. Each sensor forwards the packet with probability P_T and is silent with probability $1 - P_T$. The probability P_T is defined as a function of residual energy of the sensor. The authors proved the existence of the Nash equilibrium for this game.

In [38], a reliable routing algorithm is presented to prolong the network lifetime. Each sensor forwards the packet to its distant neighbor in order to save the total network energy by decreasing hop counts. Moreover, shorter paths are preferred for improving the reliability of transmission. Authors modeled these conflicting preferences for sensors by game theory. They demonstrated the achievability of NE. In [16], an energy-constrained routing protocol was proposed, based on a non-cooperative game. This game-theoretic model considered not only energy constraints but also path length and path reliability as a cost function for routing decisions.

In [7], Heterogeneous Balanced Data Routing (HDBR) was presented to construct energy balanced routing trees in heterogeneous WSNs. Authors modeled the game as Stackelberg game. In the game, parent sensors assume leader role and child sensors assume follower role. Parents have cooperative behavior, so they try to decrease the load of other adjacent parents. Children have selfish behavior and compete to gain more individuality. This algorithm considers bandwidth and delay in addition to the traffic load on each node as balancing criteria. Duty cycling is another effective way to reduce the energy consumption of sensors. Duty cycling mechanisms reduces the radio transmitter energy consumption of sensors in idle listening. A CSMA-based mechanism called B-MAC is developed in [23]. It utilizes low power listening and an extended preamble to obtain low energy consumption communication. Each sensor has an independent schedule. Two modes are defined for each sensor: wakeup and sleep mode. During the wakeup period, a sensor senses the channel and if a preamble is detected it remains in the wakeup mode to receive the packet. Although B-MAC performs quite well, it suffers from the overhearing problem. a MAC protocol (S-MAC) is designed for wireless sensor networks in [36]. S-MAC limits the idle listening duration by putting sensors into sleep mode periodically. Moreover, S-MAC uses message passing to reduce control overhead and latency. The protocol uses in-channel signaling, which avoids energy waste caused by unnecessary traffic overhearing. However, S-MAC suffers from the early sleeping problem [8]. In addition; S-MAC is designed to fix duty cycle that cannot be adapted to the variant traffic load [18]. However, most of the proposed MAC protocols develop a heuristic wakeup/sleep schedule, which may undermine energy efficiency and throughput performance.

Many of the proposed game theoretic approaches employ duty cycling for energy conservation. Sensors attempt to stay alive as long as possible by energy saving. Moreover, they have to provide the required services. Game theory analyzes the interaction of decision makers (i.e. sensors) to identify the conflicting interests. A game-theoretic constraint optimization scheme for duty cycle scheduling is proposed in [37]. Authors consider two parts for time: an active part and a sleeping part. During the active part, each sensor competes for the channel in the incompletely cooperative game. During the sleep part, each sensor turns off its radio transceiver to conserve its energy. The game includes many time slots and each time slot corresponds to one game state. In each time slot, each sensor estimates the current game state, based on its history. The game state is defined in term of the number of active opponents. Having estimating the game state, the sensor adjusts its own equilibrium strategy by tuning its local contention parameters (i.e. Contention Window). Then, all the nodes take actions simultaneously (i.e. transmit-ting, listening, or sleeping). Although the player does not know which action the other nodes (i.e. its opponents) are taking now, it can predict its opponents actions according to its history. Based on the estimated game state, each sensor achieves the global optima by adjusting its transmission and sleeping probability simultaneously.

Authors in [20] propose a duty cycle control mechanism for energy conservation in a sensor. A wake up and sleep scheduling strategy is implemented to conserve energy and prolong network lifetime. There is a tradeoff between packet dropping probability (due to buffer overflow) and packet blocking probability (due to the sleep mode of operation) that is modeled in wakeup and sleep scheduling strategy. The problem of determining wakeup and sleep probabilities is formulated as a bargaining game. The Nash equilibrium is used as the solution of this game to obtain the optimal wakeup and sleep probabilities.

In [1], a duty cycle control approach is proposed to reduce the idle listening duration. The authors present a mechanism as a non-cooperative game. It optimizes the sleep duration to conserve the energy of power-hungry sensors. The utility function of the game is formulated as a tradeoff between the residual energy of a sensor and the residual energy of any sensor producing packet traffic towards a sensor as sender. A set of strategies for each sensor is defined as a set of different beacon rate. The sender sensor waits for a beacon signal from the receiver before starting to transmit. Since each sender receives beacon signals from several sensors, the packets are routed on multiple paths. The authors present the NE as a solution of the game.

3. **System Model and Assumptions**

A. Network Model

The proposed method considers a WSN which consists of one sink node SN and k sensor nodes$\{n_i\}_{i=1}^{k}$ have equal capacity in terms of sensing, computation, communication, and power as Figure 1. Sensor nodes are equipped with non-rechargeable batteries, with equal initial energy level ($\forall i, IE_i = IE$). Sensors are randomly deployed and both of sensors and the sink are stationary. A flat topology is assumed for network, where all sensors play the same role (i.e., generating and forwarding packet). Each sensor can communicate with sink node directly or indirectly, depending on the distance of sensor node to the sink.

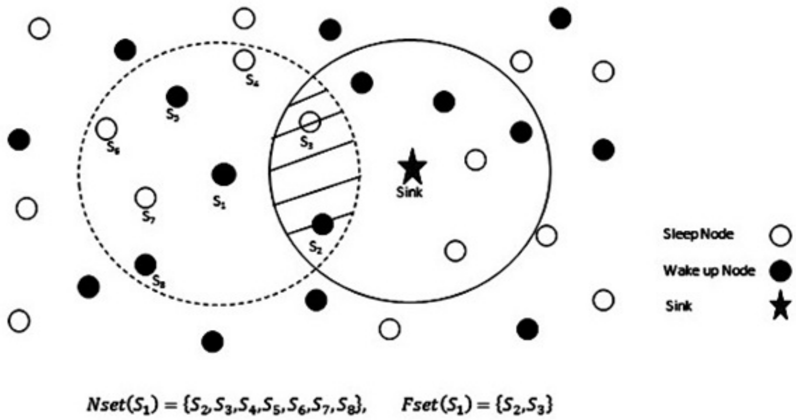

FIGURE 1. An example of neighbor set and forwarding set

B. Energy consumption model

Energy conservation is one of the most important issues in WSNs. Since sensors are powered by non-rechargeable battery, it is necessary to limit energy consumption of low power sensors to prolong the network lifetime. There are different definitions of network lifetime in the literature. We consider the network lifetime as the time until the first sensor dies [33]. In other words, the time taken till the energy of a sensor in network is depleted. The energy level of network can be defined as a tuple of all of the residual energy values of the sensors in the network, as:$< RE_1(t), RE_2(t), \ldots, RE_k(t) >$. Where $RE_i(t)$ is the residual energy of sensor i at time t that can be expressed as follows:

$$\begin{cases} RE_i(t) = AE_i(t-1) - CE_i(t-1) & if \quad t > 0 \\ RE_i(t) = IE, & if \quad t = 0 \end{cases}$$

$AE_i(t-1)$ is the available energy of sensor i at time $t-1$ and $CE_i(t-1)$ is the energy consumption of sensor i. The sensor communications usually consume more energy than sensing or processing packet. We also consider E_T as a threshold for minimum residual energy of sensors. If residual energy of a sensor is less than the E_T, the sensor will die $\exists i, RE_i < E_T$ and it is the time to compute the lifetime of the network. Network life time ratio is defined as:

$$\mathfrak{L} = \left(\frac{\min_{i \in N} (RE_i)}{\mu E_{tc} + \gamma E_{rc} + \rho E_w} \right) / E_T$$

Where E_{tc}, E_{rc} are the consuming energy for transmitting and receiving, respectively. E_w is the wasting energy [32] which contains radio amplifier circuit (E_{amp}) and idle listening (E_{idel}) as follow:

$$E_w = E_{amp} + E_{idel}$$

The network life time ratio determines how much energy remains to be used. Therefore, it can be used to evaluate effectiveness of algorithm.

C. Routing model

In this section, we propose a distributed energy aware routing algorithm for WSN.

Initialization of routing table

In the setup phase, first each sensor sends a hello packet to all neighbors that are directly connected to it. In this way, a sensor gathers its neighbors information, which contains identification, hop count to the sink, residual energy and choosing strategy. Each sensor calculates its distance to the sink in term of hop counts, drawing on neighbors information. The sensor completes its routing table. Then, it sends its routing table to all neighbors. Therefore, each sensor maintains the information of its entire neighbor in routing table to make routing decisions.

Routing decisions

Now, a fuzzy logic approach is proposed to choose the next hop in routing process. We group sensors deployed in the network according to the location of sensors. In fact, each group consists of a sensor and its neighbor nodes. In each group, the set of neighbors of sensor i that belong to G_i and are closer to the sink form a Forwarding candidate Set of sensor i named as $F_{set}(i)$. There may be overlapping in different groups. Each sensor makes routing decision independently, based on fuzzy logic. A sensor selects the next hop from Fset(i) [13] to which the packet will be forwarded. The aim is to improve the energy efficiency and prolong the network lifetime. Sensors choose wakeup or sleep strategy, based on their residual energy.

The proposed fuzzy logic algorithm makes a suitable trade-off between the energy conservation and the throughput to obtain more efficient performance. In fact, the proposed fuzzy approach considers two descriptors, namely, residual energy of a sensor and the number of awake sensors of Fset(i).Therefore, in each group only awake sensors can be selected as next hop. Hence, a sensor that has the highest residual energy among awake sensors in Fset(i) is selected as next hop according to a fuzzy logic approach. Since routing decisions are made only based on local information of sensors routing table, the new fuzzy logic approach is fully distributed. The membership functions of these parameters are depicted in Figure 2 and Figure 3.

The fuzzy if-then rules in the first level are also reported in Table 1.

Updating of routing table

Following each communication (transmission/receive packet), each sensor calculates its consuming energy and chooses wakeup strategy with probability P_{wakeup} or sleep strategy with probability P_{sleep}. A sensor chooses sleep strategy if its residual energy is equal or less than the threshold value (E_T). Sensors update their routing table information. Then, they send new information to all their neighbors. After that, all neighbors update their routing table. If neighbors do not receive any message from a sensor, it means that the sensor chooses sleep strategy.

D. Duty cycle scheduling

Duty cycle scheduling is one of the main techniques used for energy conservation in WSNs. In duty cycle scheduling approaches, each sensor frequently switches

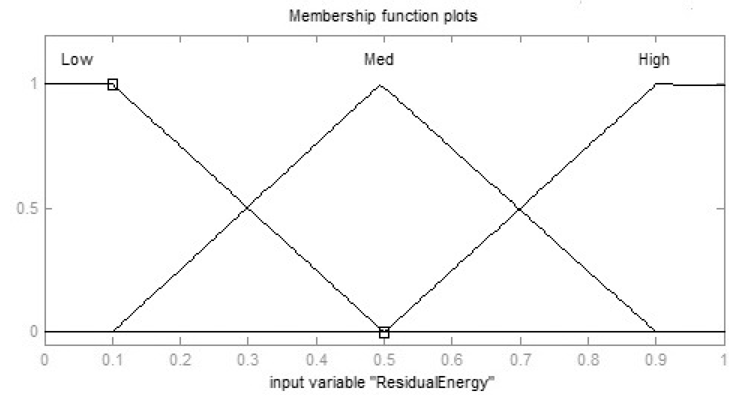

FIGURE 2. the membership function of residual energy

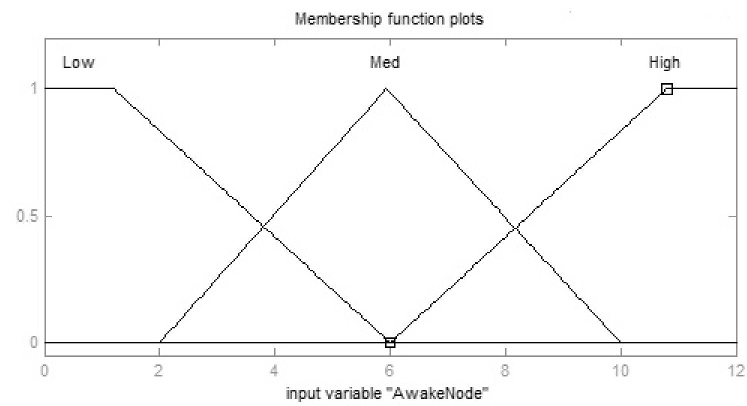

FIGURE 3. the membership function of awake node number in $F_{set}(i)$

Residual Energy	Awake node	Priority
Low	Low	Very Small
Low	Medium	Small
Low	High	Rather Small
Medium	Low	Medium Small
Medium	Medium	Medium
Medium	High	Medium Large
High	Low	Rather Large
High	Medium	Large
High	High	Very Large

TABLE 1. Fuzzy rule base

between sleeping mode and wake up mode. When a sensor is idle, it is better to turn off its radio because there is no communication in this case. Therefore, the energy can be saved by scheduling the sensor nodes efficiently. In this paper, we apply an evolutionary game theory to duty cycle scheduling for energy saving of sensors. In the wireless environments, game theory is used in order to solve many distributed resource allocation and power control problems. The evolutionary game is a powerful tool for studying and analyzing the interactions between individuals with possibly conflicting interests. We use the evolutionary game to capture the dynamic of duty cycle process. Our proposed game based approach explains how each individual get the optimal wake up scheduling and makes decision for own his benefit. A sensor changes its wake up/sleep strategy gradually when its utility is less than the average utility of all sensors. In our approach, each node decides whether or not it wakes up for the cycle to minimize energy consumption. The evolutionary equilibrium is obtained as a solution of the game. The detail of the proposed approach will be described in next section.

4. **Evolutionary Game**

A. Overview on Evolutionary Game Theory (EGT)

Recently, game theory has been successfully applied to different problems in WSNs. An EGT extends the formulation of a non-cooperative game theory by including the concept of population [21].The EGT studies the evolution of large populations of strategically interacting individuals .A population consists of a group of individuals (e.g. genes, players etc.). In the population, individuals choose different strategies simultaneously and observe their fitness (i.e. utility) while choosing their strategies. The fitness of an individual cant be measured in isolation; rather, it has to be evaluated in the context of the full population where it lives. A process similar to natural selection is used to determine how the population evolves so that the representations of the individuals having greater fitness increase in the population. According to evolutionary principles, lower fitness is the condition that causes a sub-population to shrink over time through multiple generations, eventually dying off with high probability. The EGT consists of the following two key concepts: replicator dynamics and evolutionary equilibrium. The replicator dynamics contains a set of ordinary differential equations that model process of change over time in the frequency distribution of replicators. Replicators are individuals in a population of which copies are made through the process of mutation and selection. It should be noted that, as mentioned, a replicator with a higher utility has a better chance to be copied. The EGT uses the replicator dynamics to formalize the dynamics of reproduction process. The replicator dynamics is expressed as follows:

$$\dot{x}_i = \delta x_i(\pi_i(t) - \bar{\pi}(t))$$

$x_i = \dfrac{n_i}{N}$ is the proportion of individuals choosing strategy i. n_i is the number of individuals choosing strategy i, and N is the population size. Therefore, the state of the population is denoted by $S = [x_1, \ldots, x_i, \ldots, x_G]$, the population state is a distribution x over pure strategies, where G is the total number of groups in the network, δ is the multiplier of the difference of the individuals utility and the average utility, $\pi_i(t)$ is the utility of individual choosing strategy i,and $\bar{\pi}(t)$ is the average utility of the individual:

$$\bar{\pi}(t) = \sum_{i}^{G} x_i \pi_i(t)$$

The replicator dynamics can provide information about the population (e.g., proportion of individuals who choose different strategies), given a particular point in time. The second concept of the EGT is an evolutionary equilibrium. The evolutionary equilibrium is a solution of the game. The fixed points of the replicator dynamics is the evolutionary equilibrium. Fixed points describe populations that are no longer evolving where utilities of all players are identical and no player is willing to change its strategy (to gain higher utility). The replicator dynamics also captures the speed of convergence of strategy adaptation to the evolutionary equilibrium where the rate of strategy change is zero (i.e. $\dot{x}_i = 0$).

B. Evolutionary Game Model of sensor wakeup/sleep scheduling

We propose a new distributed model for wakeup/sleep scheduling problem in WSN, based on dynamic EGT. The EGT captures the dynamics of duty cycle scheduling, based on the available local information in each sensor. In duty cycle scheduling, each sensor switches frequently between wakeup and sleep mode. Choosing an optimal duty cycle to achieve high throughput and energy conservation is the main challenge. Incoming packets will be blocked during the time a sensor is in sleep mode [20]. If sleep duration is too long, the buffer of sensor will be full. This results in dropping the packets. Therefore, it is necessary to adjust wakeup/sleep duration to reduce packet blocking/dropping probability. Moreover, there is a tradeoff between energy conservation and throughput. The evolutionary game model for wakeup/sleep scheduling is defined as:

- **Players** : the sensors which adjust their duty cycle are the players of the game.
- **Population** : the population is a set of sensors in the game.
- **Strategy** : Each player i has the two mixed strategy referred to strategy profile $Si = \{0, 1\}$, the value 0 which means sleep mode and the value 1 which means the wakeup mode. A sensor makes decision whether to receive and forward a packet to next hop (i.e. wakeup mode) or ignores the packet (i.e. sleep mode) given its residual energy. In fact a sensor chooses the wakeup strategy with probability of P_{wakeup} that can be expressed as :

$$P_{wakeup,i} = \frac{RE_i^{(g)}(t))}{\bar{E}^{(g)}(t)}$$

Where $RE_i^{(g)}$ denotes the residual energy of sensor i and $\bar{E}^{(g)}$ is the average of energy residuals of sensors in group (g) at time t. Therefore, the probability of choosing the sleep strategy can be obtained as: $P_{sleep,i} = 1 - P_{wakeup,i}$.

- **Utility** : An associated amount of benefit which each sensor in a group like g receives when it plays one of its available strategies is utility $\pi_i^{(g)}$.

$$\pi_i^{(g)}(S) = U_i(S) - C_i(S)$$

S is a vector of chosen strategies by individuals of the group g, $U_i(S)$ is a utility function to quantify a sensors satisfaction on achievable throughput. We use a concave utility function to denote the benefit of each sensor as follow:

$$U_i(S) = TR_{j,i}^{(g)}(S)$$

Where $TR_{j,i}^{(g)}$ as the throughput ratio of sensor j, which determines the proportion of sensor j of the total numbers of packets, is successfully delivered to the destination in the network.

$$TR_i^{(g)}(S) = \frac{{}^{(g)}_{j,i}(S)}{\tau_{total}} * ((1 - P_f)^{h(d)})$$

$\tau_{j,i}^{(g)}(S)$ is the throughput of a sensor j that chooses i in group (g). We define throughput to be the total number of packets that can be transferred from one node per time unit. τ_{total} is the total end-to-end traffic that gets injected into the network per time unit, P_f is the probability of transmission failure as follows:

$$P_f = P_d + P_b - P_d * P_b$$

Where P_d is the packet dropping probability and P_b is the packet blocking probability [31]. In this game, there may be no sensor to apply wakeup strategy in $F_{set}(i)$ of node i so the packet will be dropped (i.e. $P_d = 1$). If only one sensor applies wake up strategy and other sensors apply sleep strategy, the packet may be blocked with P_b probability (due to buffer overflow). This is because the awake sensor may receive multiple packets from its different neighbors simultaneously. The packet can be delivered to destination successfully, if one or more sensors select dense network, $h(d)$ can be approximated by $[\frac{d}{r}]$, (r is the transmission range of each sensors that is equal for all sensors).

$C_i(S)$ is a cost function in term of energy a sensor consumes to select strategy i.

$$C_i(S) = EC_{j,i}^{(g)}(S)$$

$EC_{j,i}^{(g)}(S)$ is energy consumption of sensor j in the group(g) to choosie strategy i that can be obtained as [32] :

$$\begin{aligned} EC_i^{(g)}(l,d) &= E_{tc,i}^{(g)}(l) + E_{rc,i}^{(g)}(l) + E_{amp,i}^{(g)}(l,d) + E_{idel,i}^{(g)} \\ &= l.E_{tran,i}^{(g)} + l.E_{recv,i}^{(g)} + l.\varepsilon_{amp}.d + E_{idel,i}^{(g)} \end{aligned}$$

Where $E^{(g)}_{tc,i}(l), E^{(g)}_{rc,i}(l)$ are the consuming of energy for transmitting and receiving l bits packet unit, respectively. $E^{(a)}_{amp,i}(l,d)$ is the energy needed by the radio amplifier circuit to send l bits packet to destination in d meters where a is path loss $(2 \leq a \leq 5), E^{(g)}_{idel,i}$ is the energy consumption for idle listening in the sleep mode, $E^{(g)}_{tran,i}, E^{(g)}_{recv,i}$ are the consuming energy for a single bit transmitting/receiving and ε_{amp} is the transceivers energy dissipation which can be expressed as:

$$\varepsilon_{amp} = \frac{\frac{S}{N}.NF_{RX}.N_0.BW.(\frac{4\pi}{\lambda})^{\gamma}}{G_{ant}.\eta.R_{bit}}$$

Where $\frac{S}{N}$ is the signal to noise ratio at the receiver, NF_{RX} is the receiver noise, N_0 is the white Gaussian noise, BW is the channel bandwidth, λ is the wavelength in meters, G_{ant} is the channel gain, η is the transmitter efficiency and R_{bit} is the data rate as bits per second.

C. Replicator Dynamics and Evolutionary Equilibrium

The EGT model is used by the duty cycle scheduling problem to reduce the energy consumption in network in WSN. The use the EGT is aimed at obtaining the optimal wakeup/sleep scheduling, resulting in a suitable trade-off between the energy conservation and the throughput in each sensor. In each iteration of the game, sensors choose a strategy from their strategy profile. As already mentioned, each sensor chooses wakeup strategy with P_{wakeup} probability. Therefore, it can receive and forward packet to the next hop. Sensors choose sleep strategy with P_{sleep} probability. In this case, they do not play any role in routing packet. However, the sensor observes its utility and compares it with the average utility of the population in the same group. Sensors will change their strategy if their obtained utility is lower than the average utility of population in the same group. Then, in the next iteration, the sensor adopts a strategy that gives a higher utility. The game is repeated, and in each iteration, the goal of each sensor is to maximize its own utility by choosing a proper strategy. Replicator dynamics captures the rate of strategy change in the population over time. The replicator dynamics of each group can be calculated as:

$$\dot{x}^g_i = \delta x^g_i(\pi^g_i - \bar{\pi}^g)$$

x^g_i is the proportion of sensors choosing the strategy i in the group g so that $\sum_i x^g_i(t) = 1, \forall t. \bar{\pi}^g$ is the average utility of the population in group g and π^g_i is the utility of a sensor choosing strategy i in the group g as:

$$\pi^{(g)}_i(S) = \frac{\tau^{(g)}_i(S)}{(\sum^G_j \tau_{j,i}} * ((1-P_f)^{h(d)}) - \sum^G_j EC_j(l,d)$$

So,

$$\sum_{j}^{G} \tau_{j,i} = n_i^G * \tau_i$$

n_i^G is the total number of sensors choosing strategy i in the entire network.

Therefore, we can obtain the replicator dynamics for our scenario as follows:

$$\begin{aligned}
\dot{x}^1_{wakeup} &= \delta x^1_{wakeup}(\pi^1_{wakeup} - \bar{\pi}^1) \\
&= (x^1_{wakeup} - (x^1_{wakeup})^2)(\pi^1_{wakeup} - \pi^1_{sleep}) \\
\pi^{(1)}_{wakeup}(S) &= \frac{\tau^{(1)}_{wakeup}(S)}{n_{wakeup} * \tau_{wakeup}} * ((1-P_f)^{h(d)}) - \sum_{j}^{N} EC_j(l,d) \\
n_{wakeup} &= N^1 * x^1_{wakeup} + N^2 * x^2_{wakeup} + N^3 * x^3_{wakeup}
\end{aligned}$$

Where N^g is the number of sensors in group g and N is the total number of sensors in the network. Since the replicator dynamics satisfies the condition of $\sum_i \dot{x}_i^g = 0$, we can calculate $\dot{x}^1_{sleep}$. The replicator dynamics of other groups ($\dot{x}^2_{wakeup}, \dot{x}^3_{wakeup}$) can be obtained in the same way. A fixed point of the replicator dynamics is an evolutionary equilibrium. In the evolutionary equilibrium situation, the rate of strategy change is zero (i.e. $\dot{x}_i^g = 0$) . In fact, in this situation, the utilities of all sensors are identical and no sensor tends to change its strategy. The evolutionary equilibrium is defined as a solution to duty cycle scheduling game.

5. Q-Learning based algorithm

In this section, we present a joint duty cycling and energy aware routing algorithm based on Q-learning. Q-learning is the most widely used and popular RL (Reinforcement Learning) approach in wireless networks [31]. Sensors use Q-learning [36] to find the optimal duty cycle, where each sensor uses only its local information. Drawing on their learning experiences, sensors make their decision (based on the local information only) independently. Therefore, it is not necessary to obtain the complete information of other sensors. Each sensor adjusts its radio duty cycle, based on its residual energy. Then, it observes its obtained utility and updates its Q-value as:

$$Q_i(t+1) = \delta(\pi_i + \varphi \max(Q_i(t))) + (1-\delta)Q_i(t)$$

Where $0 \leq \delta \leq 1$ and $0 \leq \varphi \leq 1$ are the learning rate and discount factor, respectively. Q-values represent the expected future discounted utility related to taking each action. However, in the next period, the sensor changes the radio duty cycle if necessary. In fact, the sensor chooses an action that gives a higher utility than average utility of population. The duty cycle algorithm is represented as follows (Table 2 and Table 3):

1.	**Initialize** the Q-values for all sensors in the network (Q(0))=0).
2.	**Repeat**
	a.**If** rand() $< \varepsilon$
	Choose an action (wakeup or sleep) randomly.
	"Exploration step"
	b.**Else**
	Choose the action i which maximizes the function
	$Qi(t)[i = argmaxi(Qi(t))]$. "Exploitation step"
	End if
3.	**Compute** the reward of the chosen action $\pi_i(t)]$.
4.	**Update** the Q-value as
	$Q_i(t+1) = \delta(\pi_i + \varphi \max(Q_i(t))) + (1-\delta)Q_i(t)$.
5.	**End loop**

TABLE 2. The Q-Learning Algorithm

1.	1.**If** initializing phase not be done then
	a.**For** all sensors in the network
	i.**Broadcast** Hello Packets and create routing table.
	// initialization & Phase
	b.**End loop**
	c.**For** each sensor i
	i.**Create** Neighbor set and Forwarding Set $F_{set}(i)$ of sensor i
	d.**End loop**
2.	**end if**
3.	**Choose** an awake sensor of $F_{set}(i)$ as a next hop that has the highest residual energy among other sensors and forward the packet to it
4.	**Compute** the residual energy and throughput ratio of the current sensor
	a.**if** its residual energy of the current sensor $<= E_T$
	i.**then** the sensor chooses the sleep action
	b.**end if**
5.	**Update** the routing table of the sensor and broadcast it to all neighbors

TABLE 3. Computing the reward of an action

6. Simulation and Results

In this section, we present the simulation results to evaluate the performance of proposed algorithm. We consider a network which consists of 100 sensor nodes deployed randomly in an area of 180*180 m^2 as well as a stationary sink node which is placed at $(0, 0)$ coordinate. We assume energy model based on IEEE 802.15.4 MAC layer standard. The sensor nodes are also stationary and have the same initial energy of 1J Simulation parameters are stated in Table 4.

Parameters	Value	Description
E_{recv}, E_{tran}	$50nJ/bit$	The energy consumption in receiving /transmitting a Bit
E_{idel}	$10nJ/bit$	The energy consumption in sleep mode for listening
E_{amp}	$100pJ/bit/m^2$	The energy consumption of radio amplifier circuit
E_T	0.0015	The threshold value for energy
N^1, N^2, N^3	18,35,47	the number of sensors in the groups
ρ	1	The multiplier of replicator dynamics
P_f	0	The first probability of transmission failure
P_d	0	The first probability of packet dropping
P_b	0	The first probability of packet blocking
Pactive	50%	The first probability of active mode for each sensor
Psleep	50%	The first probability of sleep mode for each sensor
R	10	The transmission range of each sensors
L	$100bits$	The packet length
ε	0.4	The probability of performingexploration step
Δ	0.2	The learning rate
φ	0.2	The discount factor

Table 4. Simulation Parameters

A. Strategy Adaption and Evolutionary Equilibrium

We present numerical results to validate the theoretical findings in this section. Figure 4 shows the phase plane of replicator dynamics. The proportion of sensors is investigated in groups G^2, with G^3 choosing wakeup strategy. We assume proportion of sensors in groups G^2, G^3 choosing wakeup strategy are $x^2_{wakeup} = 0.1$, $x^3_{wakeup} = 0.3$ as initial point. Figure 4 shows the convergence of our algorithm to the evolutionary equilibrium ($x^2_{wakeup} = 0.4$, $x^3_{wakeup} = 0.3$) under different initial points. A trajectory of energy consumption balancing shows a solution of game that approaches equilibrium. As it is shown in this figure, the evolutionary equilibrium is reached, given different initial points in less than 20 iterations.

B. Performance of Q- Learning Based Algorithm

Figure 5. demonstrates the effectiveness of our proposed joint duty cycle scheduling and energy aware routing algorithm based on Q- learning. Through learning, the sensors learn to choose wakeup/sleep action to prolong the network lifetime. Clearly, the values of two important parameters ε (i.e. exploration probability) and δ (i.e. learning rate) have a significant impact on the performance of algorithm. If the value of ε is too large (e.g.$\varepsilon = 0.9$), the algorithm will tend to make decisions randomly (i.e. exploration step). Algorithm makes decision randomly

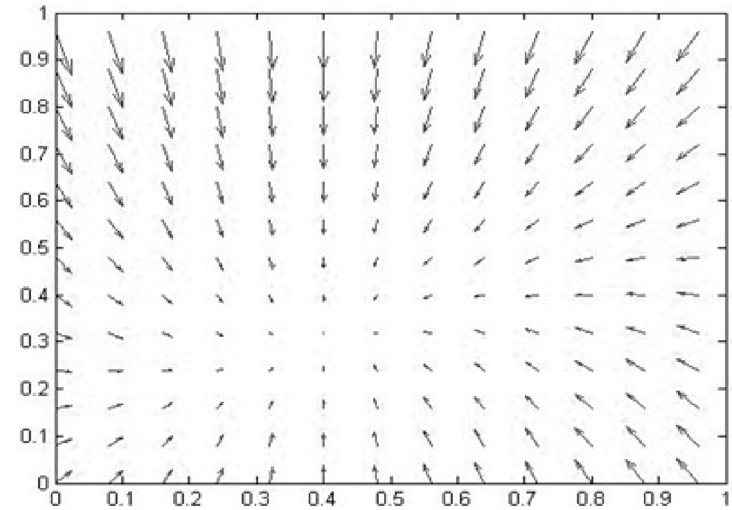

FIGURE 4. Phase plane of replicator dynamics

in order to complete knowledge in estimation of the Q-values. As the algorithm chooses actions randomly, it might not discover better actions. If the value of ε is too small (e.g. $\varepsilon = 0.1$), the algorithm is given higher priority to exploitation. Algorithm chooses the best-known action, which has the highest Q-value, in order to improve network performance. However, the algorithm has spent a little time for learning (i.e. exploration step). The knowledge is incomplete so the algorithm may be stuck in suboptimal solution. As a result, a well-balanced tradeoff between exploitation and exploration helps to maximize accumulated rewards as time goes by. However setting the value of ε appropriately balances tradeoff between exploitation and exploration, helping to achieve the desired performance of the algorithm (e.g. $varepsilon = 0.4$).Similarly, the value of δ (i.e. learning rate) plays an important role in determining the learning behavior of algorithm. Setting a large value for the δ results in faster learning (i.e. exploration phase). However, too large value of learning rate may cause fluctuations in Q-values. If the value of δ is lower, the algorithm is performed more randomly. A value of can be found so that the algorithm converge to the evolutionary equilibrium (e.g. $\delta = 0.2$).

C. Throughput

Figure 6 depicts the variations of network throughput for the proposed approach as compared to EMGT [20] and CBPA [29] for different values of wakeup probability. We define the network throughput as the total numbers of packets which are successfully delivered to the destination. The network throughput will increase if the wakeup probability becomes higher. Therefore, the network throughput changes under different values of wakeup probability. As shown in Figure 6, our approach

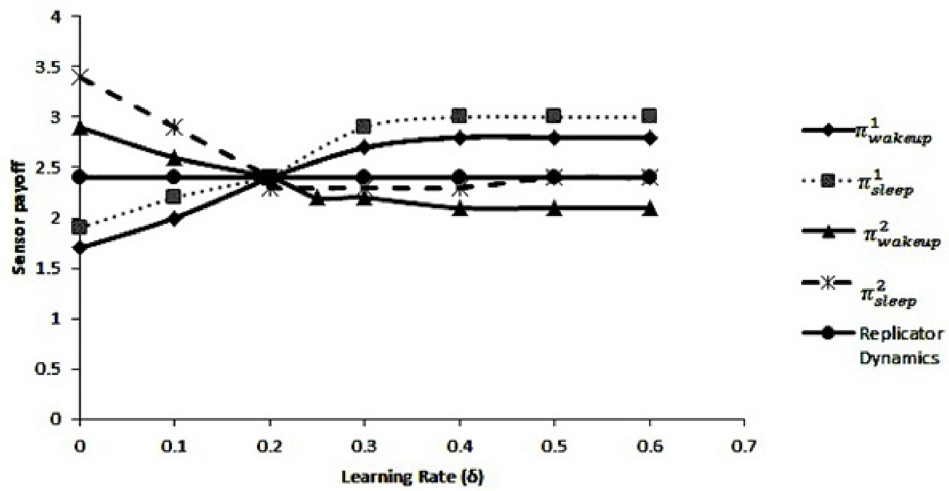

FIGURE 5. Performance of Q-learning based algorithm

achieves a higher throughput than EMGT and CBPA. The proposed approach is based on evolutionary game theory. In EGT, sensors learn how to play and choose their way gradually toward the optimal strategy. In fact, they experiment with strategies, observe their utilities, and try other strategies to obtain maximum utility.

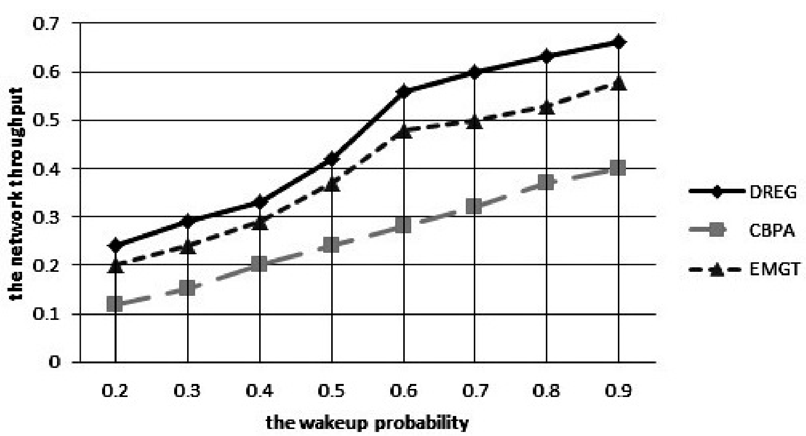

FIGURE 6. The network throughput comparison under different the wakeup probability

D. Energy consumption

Figure 7 illustrates energy consumption of proposed approach (DREG), EMGT and ENSG [17].We measure the energy consumption of sensors in many rounds. As it is shown in the figure, energy consumption increases in EMGT and ENSG if the traffic load increases. However, the proposed approach has a smaller gain than two other methods. We apply duty cycle scheduling to reduce the energy consumption in our proposed game theoretic formulation. A sensor attempts to maximize its utility by choosing an optimal strategy wakeup/sleep. As the energy consumption is smaller in sleep mode, the sensor is more preferably put in sleep mode. However, there is the tradeoff between energy conservation and Quality of Service. Simulation results show that the proposed algorithm can achieve 20% gain in energy conservation compared to the other approaches.

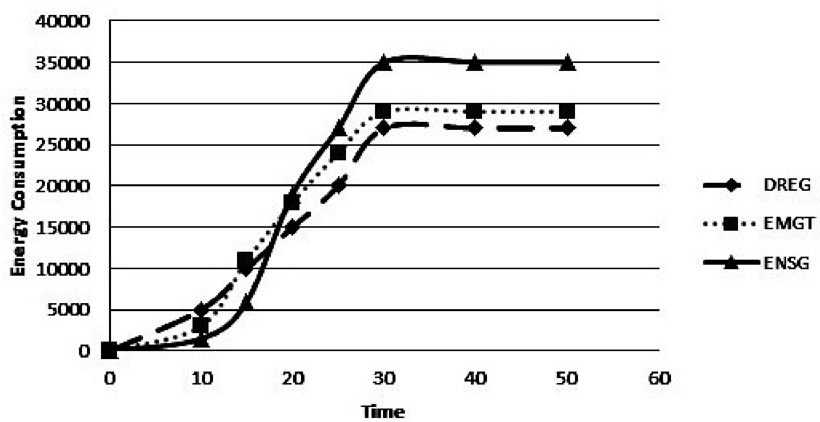

FIGURE 7. The Energy Consumption Comparison

E. Network lifetime

The proposed approach develops energy aware fuzzy based routing that considers the residual energy of nodes in routing decision. In this approach, sensors select a awake node based fuzzy rule as the next hop which has the highest residual energy among awake sensors in $F_{set}(i)$, forwarding packet to it. Therefore, the approach aims to prevent fast energy depletion of sensors with less energy and balances energy consumption through the entire network (Figure 8). Also the duty cycle scheduling employed in the proposed approach saves the energy of sensors and prolongs the network lifetime. Figure 9 shows the minimum residual energy of network until the network partitioning for the two fuzzy based algorithm (DREG) and ENSG method. The results prove that DREG performs much better than the ENSG in large networks. It is obvious from this figure that the minimum residual energy of sensor in the proposed routing protocol is higher than the ENSG algorithm. As can be seen, the proposed algorithm prolongs the death time of the first node (i.e.

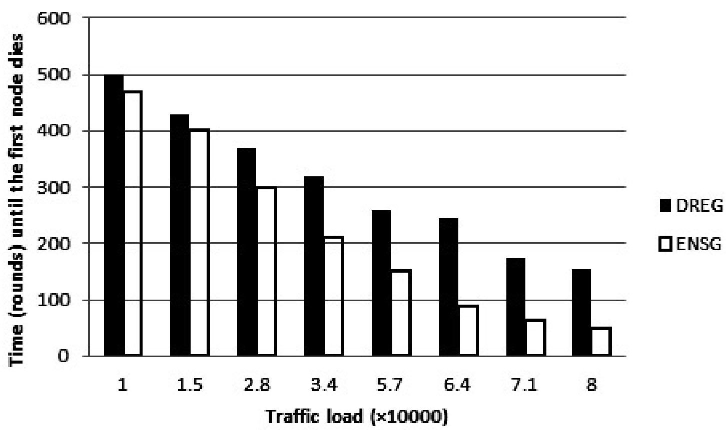

FIGURE 8. The Network Lifetime Comparison

network lifetime) in comparison with the greedy method. These results reflect the unbalanced energy consumption in the ENSG routing. Although the death time of first node decreases as traffic load increases but DREG can balance their energy consumption and avoid early energy depletion of sensors with less residual energy. The proposed fuzzy routing technique results in energy consumption balance and maximization of the overall network lifetime.

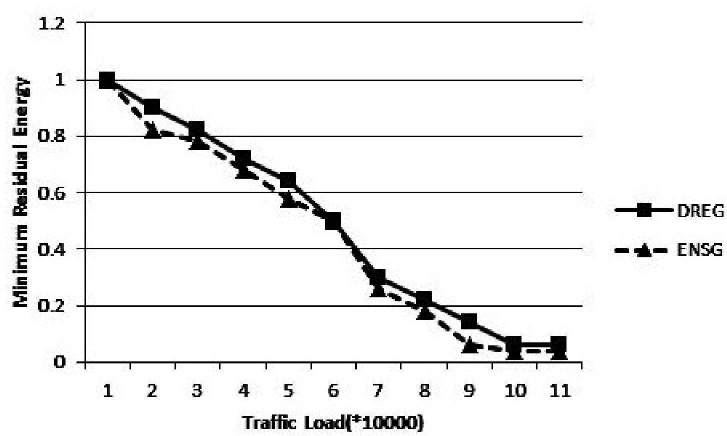

FIGURE 9. Comparison of minimum residual energy of DREG and ENSG

7. Conclusion

In this paper, a new joint duty cycle scheduling and energy aware routing approach was developed. The duty cycle scheduling problem is modeled as an evolutionary game to obtain an optimal wakeup/sleep strategy for purpose of reducing energy consumption. In the proposed approach, each sensor makes a suitable tradeoff between the energy conservation and the throughput to obtain more utility. Hence, sensors may change their strategy according to the gained utility. Based on the evolutionary equilibrium, the optimal values of the wakeup probabilities could be computed so that the throughput of a sensor node was maximized while the power supply constraints are satisfied. Moreover, we have proposed energy aware fuzzy based routing according to which a sensor with the highest residual energy among awake nodes is chosen as next hop to forward the packet. The proposed routing approach enjoys balanced energy consumption among the sensor nodes in the network, avoiding rapid energy depletion of sensors with less energy.

Finally, a Q- learning algorithm is propose for dynamic evolutionary game, based on duty cycle scheduling to obtain the evolutionary equilibrium. It is demonstrated that in the proposed Q-learning based algorithm, sensors learn and make the best decision for wakeup/sleep scheduling based only on local information. As a result, sensors improve their utility to approach the evolutionary equilibrium. The proposed evolutionary equilibrium is a state in which no sensor changes its strategy and it is defined as a solution of this game. Results of the proposed simulation study show that the Q-learning algorithm can achieve the evolutionary equilibrium. The results also show that using fuzzy logic in routing decision for next hop selection reduces the average energy consumption (by 20%), compared to energy aware routing. The results also show improvements in the performance of the system in terms of network throughput and network lifetime.

Although in the past years many game theory approaches have been proposed to address energy efficiency and lifetime maximization problems of WSNs, we propose the following topics for the future studies:

- Designing a data aggregation algorithm to reduce the number of transmitting packets and hence to reduce energy consumption
- Developing a multipath routing protocol to guarantee the reliability demand of different applications
- Designing multi-criteria (e.g., energy consumption and latency) algorithm that optimizes multiple QoS goals simultaneously to meet the demands of practical applications

References

[1] A. Abrardo, L. Balucanti and A. Mecocci, *A game theory distributed approach for energy optimization in WSNs*, ACM Trans SensNetw **9(4)** (2013), 44.

[2] I. F. Akyildiz and W. Su, Y. Sankarasubramaniam, and E. Cayirci, *Wireless sensor networks: a survey*, Elsevier Computer Networks **38**(2002b), 393-422.

[3] T. AlSkaif, M. G. Zapata and B. Bellalta, *Game theory for energy efficiency in Wireless Sensor Networks: Latest trends*, Journal of Network and Computer Applications **54**(2015), 33-61.

[4] Sh. Arafat, A. AziziMohd, N. CheeKyun, N. Nor Kamariah, S. Aduwati and Y. MohdHanif, *Review of Energy Conservation Using Duty Cycling Schemes for IEEE 802.15.4 Wireless Sensor Networks*, Wireless Personal Communications, Springer,**77**(2014), 589-604.

[5] T. Arampatzis, J. Lygeros and S. Manesis, *A Survey of Applications of Wireless Sensors and Wireless Sensor Networks*, In 13th Mediterrean Conference on Control and Automation. Limassol, Cyprus, (2005), 719-724.

[6] M. Ayers, L. Yao, *Gureen Game, An energy-efficient QoS control scheme for wireless sensor networks* In Proceedings of 2011 International Green Computing Conference, Orlando, FL, USA, (2011), 25-28.

[7] A. Behzadan and A. Anpalagan, *Prolonging network life time via nodal energy balancing in heterogeneous wireless sensor networks*, In: 2011 IEEE international conference on communications, Kyoto, Japan (2011), 1-5.

[8] M. Buettner, G. V. Yee, E. Anderson, and R. Han, *X-MAC: a short preamble MAC protocol for duty-cycled wireless sensor networks*, In Proc. of the 4th International Conference on Embedded Networked Sensor Systems , (2006), 307-320.

[9] Sh. Sh. Chiang and Ch. H. Huang, *A Minimum Hop Routing Protocol for Home Security Systems Using Wireless Sensor Networks*, IEEE Transactions on Consumer Electronics,**53(4)**(2007).

[10] A. M. Colman, *Cooperation, psychological game theory, and limitations of rationality in social interaction*, Behavioral and Brain Sciences**26**(2003), 139-198.

[11] J. C. Dagher, M. W. Marcellin and M. A. Neifeld, *A theory for maximizing the lifetime of sensor networks*, IEEE Transaction on Communications, **55(2)**(2007), 323-332.

[12] D. Fudenberg and D. K. Levine, *The Theory of Learning in Games. Cambridge*, MIT Press, Cambridge, MA, (1998).

[13] T. He, J. A. Stankovic, Ch. Lu and T. Abdelzaher, *SPEED: A Stateless Protocol for Real-Time Communication in Sensor Networks*, Proceedings of IEEE International Conference on Distributed Computing Systems, (2005), 46-55.

[14] M. Javidi and L. Aliahmadipour, *Application of game theory approaches in routing protocols for wireless networks*, In proceedings of 2011 International Conference on Numerical Analysis and Applied Mathematics, Halkidiki, Greece, (2011), 19-25.

[15] Z. Jia, M. Chundi and H. Jianbin, *Game theoretic energy balance routing in wireless sensor networks*, In Chinese control conference, (2007), 420-424.

[16] R. Kannan and S. S. Iyengar, *Game-Theoretic Models for Reliable Path-Length and Energy-Constrained Routing with Data Aggregation in Wireless Sensor Networks*, IEEE JSAC,**22(6)**(2004), 1141-1150.

[17] K. Lin, T. Xu, M. M. Hassan and A. Alamri *An Energy-efficiency Node Scheduling Game Based on Task Prediction in WSNs* , Mobile NetwAppl, Springer Science and Business Media New York, **20**(2015), 583-592.

[18] G. Lu, B. Krishnamachari and C. S. Raghavendra, *An adaptive energy-efficient and low-latency MAC for tree-based data gathering in sensor networks*, Wirel. Commun. Mob. Comput, Published online in Wiley Inter Science.**7**(2007), 863-875.

[19] R. Machado and S. Tekinay, *A survey of game theoretic approaches in wireless senso rnetworks*, ComputNetw, **52(16)** (2008), 3047-3061.

[20] D. Niyato and E. Hossain, *wireless sensor networks with energy harvesting technologies: a game-theoretic approach to optimal energy management*, IEEE Wireless Communications, (2007).

[21] D. Niyato and E. Hossain, *Dynamics of Network Selection in Heterogeneous Wireless Networks: An Evolutionary Game Approach*, IEEE Transactions on vehicular technology,**58(4)**, (2009).

[22] N. A. Pantazis, S. A. Nikolidakis and D. D. Vergados, *Energy-Efficient Routing Protocols in Wireless Sensor Networks: A Survey*, IEEE Communications Surveys & Tutorials,**15(2)**(2013).
[23] J. Polastre, J. Hill, and D. Culler, *Versatile low power media access for wireless sensor networks. In The Second ACM Conference on Embedded Networked Sensor Systems*, (2004), 95-107.
[24] O. Powell and A. Jarry, *Gradient Based Routing in Wireless Sensor Networks: a Mixed Strategy*, CoRR Distributed, Parallel and Cluster Computing, (2005).
[25] R. Rajagopalan and P. K. Varshney, *Data aggregation techniques in sensor networks: A survey*, IEEE Commun. Surv. Tutor.,**8**(2006).
[26] T. Rault, A. Bouabdallah and Y. Challal, *Energy-efficiency in wireless sensor networks: a top-down review approach*, ComputNetw, **67**(2014), 104-122.
[27] H. Ren and M. Meng, *Game-theoretic modeling of joint topology control and power scheduling for wireless heterogeneous sensor networks*, IEEE Trans. Autom. Sci. Eng.**6**(2009), 610-625.
[28] A. Schillings and K. Yang, *VGTR A collaborative, energy and information aware routing algorithm for wireless sensor networks through the use of game theory*, In Proceedings of 3rd International Geosensor Networks Conference, Oxford, UK, (2009), 13-14.
[29] H. Shpungin and Z. Li *Throughput and Energy Efficiency in Wireless AdHoc Networks with Gaussian Channels*, IEEE Communications Society, (2010), 289-298.
[30] J. M. Smith, *Evolution and the Theory of Games: In situations characterized byconict of interest, the best strategy to adopt depends on what others are doing*, American Scientist, (1976).
[31] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning)*, MIT Press, Cambridge, MA, (1998).
[32] D. Tudose, L. Gheorghe and N. T. Apus, *Radio transceiver consumption modeling for multihop wireless sensor networks*, UPB Scientific Bulletin, Series C,**75(1)**(2013), 17-26.
[33] Y. Wu, Zh. Mao and S. Fahmy, *Constructing Maximum-Lifetime Data-Gathering Forests in Sensor Networks*, IEEE/ACM Transactions on Networking ,**18(5)**, (2010).
[34] G. Yang and G. Zhang, *A power control algorithm based on non-cooperative game for wireless sensor networks*, In Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology, Harbin, China, (2011), 12-14.
[35] KLA. Yau, P. Komisarczuk and P. D. Teal, *Reinforcement learning for context awareness and intelligence in wireless networks: review, new features and open issues*, J Netw Comput Appl. **35(1)**(2012), 253-267.
[36] W. Ye, J. Heidemann, and D. Estrin, *Medium access control with coordinated, adaptive sleeping for wireless sensor networks*, ACM Transactions on Networking,**12(3)**(2004).
[37] L. Zhao, L. Guo, L. Cong and H. Zhang, *An energy-efficient MAC protocol for WSNs: game-theoretic constraint optimization with multiple objectives*, WirelSensNetw, (2009), 358-364.
[38] M. Zheng, *Game theory used for reliable routing modeling in wireless sensor networks*, In International conference on parallel and distributed computing, applications and technologies, China, (2010), 280-284.

M.S. Kordafshari, Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
E-mail address: `kordafshari@qiau.ac.ir`

A. Movaghar*, Department of Computer Engineering, Sharif University of Technology, Tehran, Iran
E-mail address: `movaghar@sharif.edu`

M.R. Meybodi, Computer Engineering and Information Technology Department, Amirkabir University of Technology, Tehran, Iran
E-mail address: `mmeybodi@aut.ac.ir`

*Corresponding author