

ELACCA: Efficient Learning Automata Based Cell Clustering Algorithm for Wireless Sensor Networks

Neeraj Kumar · Jongsung Kim

Published online: 16 June 2013
© Springer Science+Business Media New York 2013

Abstract Wireless Sensor Networks (WSNs) are a special type of networks deployed in different geographical regions for capturing the important information. WSNs consist of low energy devices called Sensor Nodes (SNs) which are capable of sensing and transferring the gathered information to remote controller called as Base Stations (BSs). Because these devices are generally deployed in unattended environment and are limited in communication and computing power, so it is not always possible to recharge or replace the batteries for these devices. The SNs are supposed to have self healing and built in intelligence to operate independently. Keeping view of the above, in this paper, we propose a new Efficient Learning Automata based Cell Clustering Algorithm (ELACCA) for WSNs. Compared to the earlier approaches, we have taken size of the cell of the area under investigation in rhombus shape rather than the square. The selection of cluster head (CH) is performed by different levels using the participation ratio of the nodes in respective CH. Using the defined participation ratio, a cut off on number of nodes in a particular CH is also computed. Moreover, by varying the angle from base of the cell to its sides, the numbers of CHs formed are also calculated. Using these values, the communication among different CHs is maintained. The performance of the proposed scheme is validated using the extensive simulation with respect to various parameters such as connectivity, coverage and packet delivery ratio. The results obtained show that the proposed scheme is better than the existing schemes with respect to these metrics.

Keywords WSNs · Clustering · Learning automaton

N. Kumar
Department of Computer Science and Engineering,
Thapar University, Patiala, Punjab, India
e-mail: nehra04@yahoo.co.in

J. Kim (✉)
Department of Mathematics, Kookmin University, 77, Jeongneung-Ro,
Seongbuk-Gu, Seoul 136-702, Republic of Korea
e-mail: jskim@kookmin.ac.kr

1 Introduction

Over the past few decades, Wireless sensor networks (WSNs) have been used in many of indoor and outdoor applications. Although the size of the Sensor nodes (SNs) is small but these are capable of performing many operations in unattended geographical regions. WSNs contain nodes having limited energy and computing power deployed in a certain region for specific applications. SNs in typical WSN communicate with each other in multi hop manner for information dissemination.

With the rapid advancement of wireless technology, WSNs are now used for various mission critical applications, e.g., weather forecasting, remote healthcare and disaster management military applications. All these applications require efficient mechanisms for coverage and connectivity preservation along with energy saving as energy level of these devices deplete quickly when they perform certain operations. Specifically, the energy of the nodes is wasted to large amount if they are performing no operations in a given time slot. This may happen because nodes did not know when to transmit or receive the data from other SNs in the network [1,2]. Generally, the data collected by the SNs is transferred in hop by hop manner to the remote locations where Base Station (BS) is deployed. As this involves several round of communication among the nodes in the network so lot of energy is wasted in this mechanism. As energy is a scarce resource in WSNs so a novel mechanism is required which has self healing features so that the valuable resource such as energy can be saved.

In literature, clustering is the most common technique which is used for efficient operations to save energy in WSNs [3–10]. The area under investigation is divided into various clusters depending upon the various factors such as density, remaining energy, or coverage and connectivity. Out of the participating nodes in the network, some of the nodes may be elected as cluster heads (CHs) and rest of them are treated as non CHs. The CHs are responsible for monitoring and data transmission to BS in a particular region. All non CHs nodes transmit the gathered information to their respective CH. This process will save lot of energy as all the nodes need not remain in contact with the BS all the time. The respective CH is responsible for collecting and maintaining the data in the region under investigation.

Although clustering has been widely used in WSNs to save the energy as discussed above [3–10], but it also raises many challenges such as selection of CH nodes and coverage and connectivity in the network depending upon various metrics. An efficient coverage mechanism covers all the nodes in the network with minimum overlap and resource consumption. This will minimize the number of CHs in the network and energy consumption of the nodes. Both of these issues are challenging and required a specific approach to solve the same. There are many proposals for efficient clustering with coverage and connectivity in WSNs covering these aspects. Some protocols have used duty cycle which performs the on/off in WSNs nodes to save the energy [2]. But still there is a requirement of efficient optimized clustering mechanism which not only performs the efficient clustering and maintains the coverage and connectivity by consuming less energy but also have features of self healing and learning in it. Keeping in view of the above, in the paper we propose a new Efficient Learning Automata (LA) based Cell Clustering Algorithm (ELACCA).

It has been found in literature that LA can be used to solve wide variety of applications [11–16]. It is an optimization technique that uses the concept of learning based upon the input parameters and produces an output, i.e., it is an adaptive learning technique with a decision-making machine which has the capability to improve by learning from its environment so that it can choose the optimal action. The automaton is assumed to be deployed at each of the CH for capturing the information from the environment and then adaptively selects the operation to be performed. For each action taken by the automaton, it receives reward

or penalty and based upon these inputs from the environment, automaton decides its next action to be taken. The automaton deployed at each of SN collaborate with each other also o share the valuable information with each other. After finite number of steps, the automaton converges to particular value which is taken as the solution to a particular problem.

The rest of the paper is organized as follows. Section 2 describes the related work. Section 3 describes the background of learning automata and basic terminology used. Section 4 describes the system model and problem definition. The proposed LA based cell clustering algorithm is described in Sect. 5. Section 6 presents the simulation environment with results and discussions. Finally, Sect. 7 concludes the paper and provides the future scope.

2 Related Work

As WSNs have been used in many of our day to day life applications, lot of research proposals covering various aspects of the same have been reported in the literature. Fuad et al. [6] have proposed an energy efficient adaptive algorithm for selection of CH based upon the residual energy of the nodes. Extensive simulation with respect to various parameters is done in the proposed scheme for validation. Yong et al. [7] have proposed a new clustering algorithm which depends upon the density and size of the network. The proposed scheme is found to be better in comparison to the existing schemes. Khalil et al. [8] have proposed energy aware evolutionary algorithm in WSNs. Authors have formulated a new objective function having parameters such as network lifetime, stability and energy consumption. Jin et al. [9] have explored the coverage and connectivity problems in WSNs. Authors have assumed a uniform distribution of SNs in a circular region and then a probabilistic lower bound is derived to get an idea about the connectivity preservation for the area under investigation and an algorithm is also proposed for the same. Lin et al. [10] have proposed an ant colony algorithm for energy optimization in WSNs. In the proposed scheme, by estimating the remaining energy the selection of next hop is done in probabilistic manner. Moreover, the updates in the values of the pheromone are also done using the global and local values.

Liu et al. [17] have proposed a unique identification for every SN in the network and then provide an analytical model for CH election in distributed manner based upon the remaining energy and node distribution in a particular region of the network. Aioffi et al. [18] have proposed a multi objective network model for energy saving and increase the lifetime of the network. Authors proposed various policies for the minimization of energy consumption and data dissemination latency. Lai et al. [19] have proposed a new mechanism for energy saving by varying the size of the cluster. The reason for varying the size the cluster is that the nodes which are closer to BS consume more energy than the other nodes in the network as they suffer heavy traffic load during data dissemination from other SN to BS. Francesco et al. [20] have proposed a multi objective approach for lossy compression which gives the flexibility to the users to select the parameters depending upon a particular application. The proposed approach is tested with respect to complexity and data compression by collecting data from real environment. Ting et al. [21] have presented an efficient algorithm for increasing the lifetime of WSNs. The proposed scheme consists of exploration and at both local and global levels. The proposed algorithm is executed without any restriction on the cover set of the number of SNs in the network.

Bandyopadhyay et al. [22] have proposed a distributed clustering algorithm for energy conservation in WSNs. By arranging the CHs at different levels, authors proved that with an increase in levels of CHs, more energy can be saved. Wang et al. [23] have proposed residual energy based traffic splitting algorithm. In the proposed scheme, authors have used linear

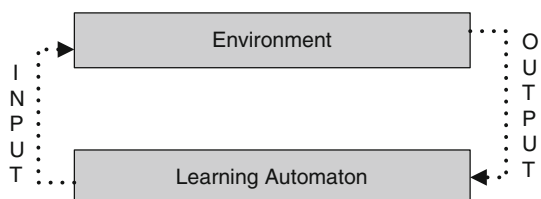
programming to describe a new protocol which distributes the energy among the participating nodes equally to prolong the network lifetime. Alippi et al. [24] have proposed an algorithm for k-level clustering by taking into consideration of node's residual energy and aggregation degree and provide a detailed simulation of the proposed scheme. Attea et al. [25] have formulated a new fitness function having parameters such as cohesion and separation error for CH election using evolutionary algorithm. The proposed algorithm tries to remove the deficiencies in the earlier algorithms by improving the network lifetime, and preservation of energy. Yu et al. [26] proposed an energy-aware clustering (EADC) routing by introducing the concept of load balancing on CHs. The proposed algorithm uses the competition range for efficient selection of CH and routing in sparse dense regions. Pinaki et al. [27] have proposed security enhanced communication in WSNs using R-M codes. Pan et al. [28] have proposed recommendation of by clustering of tag neighbors. Silas et al. [29] have proposed a novel fault tolerance service selection framework for pervasive computing. Dhurandher et al. [30] have proposed communication mechanism for underwater networks.

All of the above proposals have used various techniques for different levels of clustering to save the energy without any learning and self healing mechanisms at SNs. But in literature, it has been observed that LA based techniques are very useful in finding the optimal solution with in the specified constraints and converge rapidly to solve many engineering applications [11–16]. Esnaashari et al. [11] have proposed a strategy which guides the movements of SNs within the area of the network without any sensor to know its position or its relative distance to other sensors. Same authors [12] have proposed LA based scheduling algorithm to solve the network coverage problem in WSNs. In the proposed scheme, automaton learns the sleeping time of the node by knowing exactly the movement of the target points. Torkestani et al. [13] have described the mobility pattern of the nodes using LA by knowing the mobility pattern which can also be used in the construction of multicast tree. Also they have proposed LA based sampling algorithm to solve the minimum spanning tree problem in stochastic graphs [14]. In the proposed scheme, a set of LA are randomly activated and determine the graph edges used in sampling phases. Using this idea, authors have constructed backbone formation algorithm to find an optimal solution for the minimum connecting dominating set problem [15]. Savvas et al. [16] proposed a new method for visual multimedia content encryption using cellular automata. The scheme was based on the retrieval of the contents based upon type of application.

3 Overview of Learning Automata

Learning automata (LA) is an optimization technique which operates according to some inputs and produce the output. Figure 1 describes the generalized architecture of LA. An automaton takes some input parameters and then performs an action to produce an output. Moreover, the machine has the capability of learning from its environment so that it can choose the desired action from a finite set of allowed actions through repeated steps [11–16].

Fig. 1 Learning automata



The initial action chosen is random in nature, but by taking input from the environment in terms of reinforcement signal, after finite number of steps, the solution converges and produces the desired output.

Mathematically, LA is defined as $\langle Q, K, P, \delta, \omega \rangle$, where $Q = \{q_1, q_2, \dots, q_n\}$ are the finite set of states of LA, $K = \{k_1, k_2, \dots, k_n\}$ are the finite set of actions performed by the LA, $P = \{p_1, p_2, \dots, p_n\}$ are the finite set of response received from the environment, and $\delta : Q \times P \rightarrow Q$ maps the current state and input from the environment to the next state of the automaton and ω is a function which maps the current state and response from the environment to the state of the automaton [11–16,31,32].

The environment in which the automaton operates can be defined as a triplet $\langle X, Y, \rho \rangle$, where $X = (X_1, X_2, \dots, X_n)$ are finite number of inputs, $Y = (Y_1, Y_2, \dots, Y_n)$ are values of reinforcement signal, and $\rho = (\rho_1, \rho_2, \dots, \rho_n)$ are penalty probability associated with each $X_i, 1 \leq i \leq n$. Based upon each action taken by the automaton its action is rewarded or penalized and its action probability vector is updated as follows:

The formula for probability update is as follows [11–16]:

$$\begin{aligned}
 p_j(n + 1) &= (1 - a)p_j(n), & j \neq i, & Y = 0 \\
 p_j(n + 1) &= ap_j(n), & j = i, & Y = 0 \\
 p_j(n + 1) &= p_j(n), & & Y = 1,
 \end{aligned}
 \tag{1}$$

where a is the parameter.

4 System Model

4.1 Network Model

Figure 2 shows the network model used in the proposed scheme. All SNs are uniformly distributed in geographical region and send the data after collecting at regular interval to BS via CHs. The area under investigation is divided into various cells and each cell has cluster and sensor nodes deployed. The SNs periodically sense the data in the respective regions and send it to CHs of that region. Each CH in the respective region sends the collected information to the nearest BSs for further processing. Each CH communicates with the other CH for information sharing. Let $S = \{S_1, S_2, S_3, \dots, S_n\}$ and $CH = \{CH_1, CH_2, \dots, CH_n\}$ are the n number of SNs and CH nodes deployed in the region under investigation. We have assumed that finite number of nodes are divided into different CHs so that a CH in a particular region is overall responsible for all the activities such as data transmission, message transmission etc. Initially

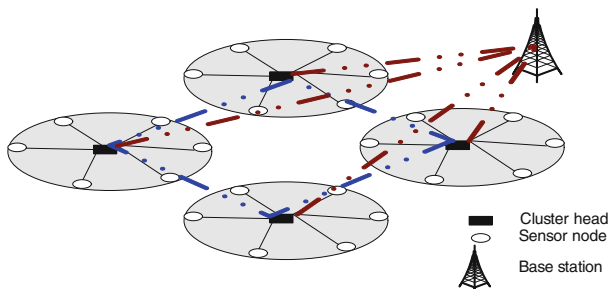


Fig. 2 Network model in the proposed scheme

all the nodes are assumed to have same energy value but as they perform any operation their energy level decreases.

To preserve the energy at various levels, we have defined an upper threshold on the value of SNs in a particular CH. This value is called as the cutoff value and is calculated by defining a function. We assume that $value(f^{cut})$ is the upper threshold on the number of nodes in a particular CH in a network field of area $[0, L]^2$ where L is the length of the cell. The cutoff for the number of nodes in a particular CH consists of parameters such as remaining energy and angular position of nodes from the base line which defines the location of the node. Also, the defined cutoff is assumed to have success or failure and accordingly its probability is also defined. This probability is calculated by varying the angular position of the nodes and remaining energy level after operating any specific operation. In this way by applying the concept of cutoff on the sensor nodes, we can consider them as heterogeneous in terms of different resources they have.

4.2 Problem Definition

Definition A cutoff is a function that defines the upper threshold on the number of nodes in a particular CH and can be defined as follows:

$$f^{cut} : K \rightarrow N^{CH} \tag{2}$$

where $K = \sum_{i=1}^n S_i$ where S_i are the SNs in a particular region and $value(f^{cut})$ is the value of the cut off nodes in a particular CH. The values in the variable $value(f^{cut})$ can be set according to the application to application. The cut off defined on the number of SNs can be successful it may be a failure. The success or failure is controlled by the blocking probability of the cutoff, i.e., the number of nodes in a particular CH should be optimized so that a particular CH should not be overloaded or under loaded. The success or failure is counted by the fact that some of nodes are not in the range of the BS or CH due to location of the same in a particular geographical region. We have taken this factor as the SNs may be deployed in dense or sparse region of any geographical location. Hence keeping in view of the same, following terms are defined as follows:

Define the probability of failure (P^F) and success (P^S) of cutoff on number of nodes in a cell as:

$$P^F = B^i \sum_{j=0}^{i-1} (1 - B^j) \tag{3}$$

$$P^S = \sum_{i=0}^n (1 - B^i) \tag{4}$$

B^i is the blocking probability of the cutoff in the region. It is defined as the ratio of upper threshold on the number of cutoff to the total number of SNs in the network.

So the combined probability of the cutoff can be defined as follows:

$$P^{com} = LP^S + \sum_{j=1}^L jP^F = L \sum_{i=0}^n (1 - B^i) + B^i \sum_{j=0}^{i-1} (1 - B^j)j \tag{5}$$

The cutoff on number of SNs is an iterative process in which after finite number of steps, the solution would converge to a particular value. The process of cutoff has some probability of

success or failure in each step as defined above. Let z are the number of steps to perform the final cutoff which includes the count of success or failure in a particular time interval.

So, cumulative (tot) probability of cutoff can be defined as follows:

$$tot = \sum_{z=1}^{\infty} z P^{com} p_s (1 - p_s)^{z-1} \tag{6}$$

Let $D(S)$ and $\overline{D(S)}$ is the amount of data disseminated before and after the cutoff operation on number of SN in particular CH. The cumulative data disseminated can be calculated in modified form as defined in [33]:

$$\overline{D(S)} = \max(D(S), D(S_1))P^S + \min(D(S), D(S_1))P^F \tag{7}$$

Equation (7) gives detailed description about the data dissemination in between the two SNs in the network before and after cutoff operation. While selecting the maximum value, probability of success of cutoff is multiplied but probability of failure during cutoff is multiplied for selection of minimum value.

Now, using Eqs. (2)–(7), we have defined a combined objective function μ^{obj} for the given problem as follows:

$$\mu^{obj} = \min\{\overline{D(S)} \times tot\} \tag{8}$$

5 Proposed Approach

In this section, we would explain various steps of the proposed scheme in detail. Basically we have assumed that each automaton is deployed at each of CH which is responsible for overall monitoring and control of the SNs in the respective regions. The operations controlled by the automaton are classified in the following steps and accordingly automaton updates its action probability vector.

5.1 Maintenance of the Connectivity in the Cell by the Automaton

Each automaton controls the connectivity ratio of the SNs for respective cells and facilitates the various operations. The connectivity ratio is the probabilistic measure of number of SNs in a particular cell to the Lot of energy is preserved by the way the particular operation is performed by the automaton. For each action taken by the automaton, its action probability vector is updated according to the input provided by the environment which is defined in the section above. The main concern in WSNs is to preserve the connectivity of the SNs. If few of the nodes are not connected to the network then communication between the SNs can not be performed and these nodes are considered to be dead nodes. The nodes in the cell are monitored by the respective CH which maintain the connectivity both in the intra and inter domain. The automaton takes input as the structure of the cell and then calculates the probability of connectivity of the nodes with respect to the position of the nodes. The position of each node in the cell is estimated by variation of the angle of inclination from the base of the cell. The area under investigation is divided in to cells with size of each cell is assumed to be as rhombus which is a variation of the assumption made in Hybrid Energy Efficient Distributed (HEED) Clustering protocol [37].

Assumptions of our protocol are similar to those for HEED [37]. The main difference between HEED [37] and the proposed scheme is the size of the cell. In the proposed scheme we have taken a rhombus shape of the cell, area ABCD is the cell structure for the sample calculation in which we have taken $\theta = 60^\circ$.

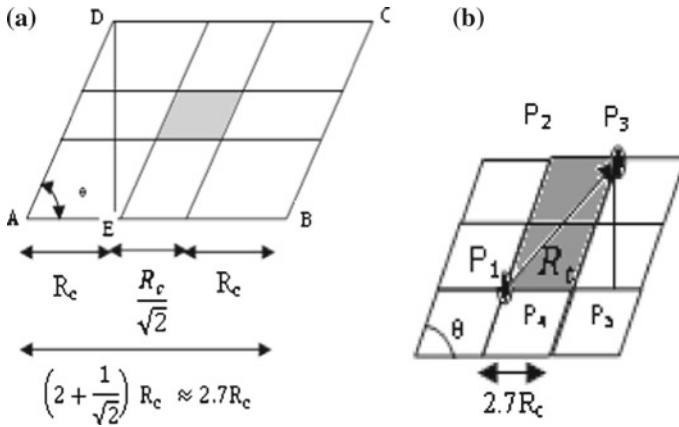


Fig. 3 a Structure of cell, b intra cluster communication

In worst case the distances between the CHs which are on the edge of cell are $2.7 R_c$ [33]. From Fig. 3, in triangle ADE,

$$\sin 60 = \frac{DE}{AD}, \text{ where } AD = \left(2 + \frac{1}{\sqrt{2}}\right) R_c \cong 2.7R_c \tag{9}$$

$$DE = AD \sin 60^\circ, \quad DE = \left(2 + \frac{1}{\sqrt{2}}\right) \times \frac{\sqrt{3}}{2} R_c \cong 2.44R_c \tag{10}$$

Therefore the area of rhombus ABCD can be easily calculated,

$$\text{Area} = AB \times DE = 2.7R_c \times 2.44R_c \cong 6.588R_c^2 \tag{11}$$

In Fig. 3, area $P_1P_2P_3P_4$ covers two cells assuming that the two CHs are located at the P_1P_3 . So even if this is the worst case that the two CHs are located at the corners of the cell, for multi hop clustering these two CHs should be connected and the distance between these two CHs will be R_t . [33]. If ‘d’ be the perpendicular distance from P_2 to P_1P_4 than we can calculate d:

$$\sin 60 = \frac{d}{P_1P_2} \text{ where } P_1P_2 = 2 \times 2.7 \times R_c \cong 5.4R_c \tag{12}$$

$d = 4.68 R_c$ further if the distance from the P_1 to the perpendicular from P_2 is ‘r’ then by simple calculation it is found that $r = 2.7 R_c$. Once these values are calculated by the automaton for respective cell then it is able to preserve the connectivity of the nodes located at different positions of the cell.

5.2 Calculation of Distance Between CHs by the Automaton

The distance between the two CHs and average number of CHs in the whole geographical region is computed next by the automaton. It keeps track of the nodes and CHs association at different time intervals with respect to the various conditions of connectivity. The distance between different CHs and number of CHs connected are required to be computed for complete operations which are calculated as follows:

- (i) Calculation for R_t :
From Fig. 2, in triangle $P_1P_3P_5$

Table 1 Variation of R_t in terms of R_c and CH selected [33]

θ	R_t in terms of R_c	No. of CHs for selected network field
90	$R_t \geq 6.000 R_c$	$0.137 \frac{L^2}{R_c^2}$
80	$R_t \geq 6.450 R_c$	$0.139 \frac{L^2}{R_c^2}$
70	$R_t \geq 6.810 R_c$	$0.146 \frac{L^2}{R_c^2}$
60	$R_t \geq 7.15 R_c$	$0.158 \frac{L^2}{R_c^2}$
50	$R_t \geq 7.42 R_c$	$0.179 \frac{L^2}{R_c^2}$
40	$R_t \geq 7.67 R_c$	$0.213 \frac{L^2}{R_c^2}$
30	$R_t \geq 7.858 R_c$	$0.275 \frac{L^2}{R_c^2}$
20	$R_t \geq 7.990 R_c$	$0.403 \frac{L^2}{R_c^2}$
10	$R_t \geq 8.070 R_c$	$0.790 \frac{L^2}{R_c^2}$

$$\begin{aligned}
 P_3P_5 &= d = 4.68R_c \\
 P_1P_5 &= P_1P_4 + P_4P_5 = 2.7R_c + 2.7R_c = 5.4R_c \\
 R_t &= \sqrt{(P_1P_5)^2 + (P_3P_5)^2} \cong 7.15R_c \tag{13}
 \end{aligned}$$

From this result it is clear that CHs are connected if $R_t \geq 7.15R_c$ when $\theta = 60^\circ$

(ii) Calculation of no. of CHs in network field:

For $\theta = 60^\circ$, we calculated the area of cell $6.588R_c^2$.

Hence numbers of CHs in the network field are

$$N^{CH} = \frac{L^2}{6.588R_c^2} \cong 0.152 \frac{L^2}{R_c^2} \tag{14}$$

Equations (13) and (14) are computed at each of the CH by respective automaton for calculation of the above defined parameters. Table 1 below computes the distance between the CHs and average number of CHs for different values of θ . Table 1 is maintained by the respective automaton at each of the CH to update these two values [33]. For each value of θ , the different conditions for maintaining the connectivity among the SNs and CHs are maintained at respective CH, i.e., by variation of θ .

Theorem *The cutoff on the number of nodes in the cluster head is $\sqrt{K} \left(\frac{L^2}{R_c^2} + 1 \right)$.*

Proof Initially we have assumed the degree of the cluster head node as m . The new node leave or join the network and can migrate from one CH to another. In this case the size of CHs change and their degree also vary. As the degree of the CH reaches a threshold value (thr) then all the nodes calculates the following values and decide their participation degree S^{part} in respective CH as follows:

$$S^{part} = \left\langle \begin{cases} N^{CH} \left(\frac{P_m}{P} \right) \right\rangle, & m \leq thr \\ \frac{value(f^{cut}) - S(f^{cut})}{S}, & m > thr \end{cases} \tag{15}$$

At any instant the position of the SNs is defined by it location L that consists of horizontal and vertical components with its degree of participation as defined above. As defined in the above

equations, degree of participation consisting of cutoff value and number of CHs within and above the defined threshold value. So the combined attributes can be defined by the following equations:

$$\begin{aligned}
 x((t + \Delta t), S^{part}) &= x(t) + x\Delta t + S^{part} \\
 y((t + \Delta t), S^{part}) &= y(t) + y\Delta t + S^{part} \\
 f^{cut}((t + \Delta t), S^{part}) &= f^{cut}(t) + f^{cut}(\Delta t) + S^{part}
 \end{aligned}
 \tag{16}$$

Subtracting the x and y co-ordinates, we get the exact position along with the participation of the nodes in the respective cluster as follows

$$L = x((t + \Delta t), S^{part}) - y((t + \Delta t), S^{part}) \tag{17}$$

$$L^2 = (x((t + \Delta t), S^{part}) - y((t + \Delta t), S^{part}))^2 \tag{18}$$

Equation (18) gives the location with respect to the participation degree and cutoff bound in the respective CH for SNs.

Dividing the value by square of the radius value R_c and multiplying by K, we get the value of cutoff in the respective CH. Hence the proof of the theorem. \square

5.3 Update of Action Probability Vector According to the Participation Degree

The main part of LA is how the action probability vector is updated. The action probability vector contains the values that needs to be updated corresponding to each action of the automaton. The environment provides a feedback corresponding to each action of the automaton. The feedback may be positive or negative and accordingly the values in the action probability vectors are updated. The participation degree of each node in a particular CH is computed by the automaton. The participation degree of the node is computed by calculating the number of packets successfully sent to the total number of packets sent with respect to the geographical region of the nodes in a particular time interval in the given network. The location is introduced in the participation degree as some of the nodes may have poor connectivity in a particular region. Moreover, the time is also a crucial factor as the information may be sent in strict deadline as defined above in Eqs. (15)–(18).

For each operation, if environment provides reward to LA then participation degree of SNs is get incremented as follows: $S^{part} = S^{part} + p_j(n)$. Similarly if environment provides penalty to LA then its participation degree of SNs is get decremented as follows: $S^{part} = S^{part} - p_j(n)$. The probability factor is changed only in the case of reward function else it remains same as we have considered the reward inactive scheme in the current proposal. Now based upon these values, the probability of success or failure and combined probability of cutoff on the number of nodes in a particular cell are computed using Eqs. (3)–(6). The cut off values are used to decide the connectivity of the nodes in a cell as illustrated above. Table 1 below provides a complete detailed overview of the various parameters variation in the proposed scheme, e.g., with an increase in the values of θ , R_t in terms of R_c is decreased and number of CHs for the selected network is also decreased.

The proposed algorithm consists of two steps: Cluster head maintenance and connectivity preservation. Cluster head maintenance consists of CH election by application of cutoff at different levels and by computing the node participation ratio in the cell. The connectivity preservation phase consists of maintaining the connectivity of the nodes with respect to their position, e.g., some of the nodes are in the middle of the cell and some of them are on the edge of the cell. In the proposed algorithm, the automaton is activated at various levels of clustering hierarchy. Initially the node having maximum energy is selected as CH and

automaton is activated at this node (lines 1–3). Once the automaton is activated at respective CH then it performs its action by computing the participation degree and also compute the probability of selection of CH for the next round (lines 4–8). The probability of success or failure is computed next (line 9). If LA receives an award or penalty the participation ratio of SNs is incremented or decremented (lines 10–15). Based upon these values, the cumulative probability of cutoff is computed using Eq. (6) (line 16). First level cutoff for SNs is applied and if participation ratio of the node is less than the value of cutoff then this node is made CH for the next level else the second level cutoff is applied to select the node for CH (lines 17–22). To preserve the connectivity of the nodes in the network, cell radius is computed using Eqs. (9), (10) and (11) and distance between them is computed if they are on the edge of the cell (lines 23–25). The angular position of the nodes from the base position is computed to find the location of the same using Eqs. (16), (17) and (18) (lines 26–32). The value of R_t is computed in terms of R_c (lines 33–35). In the final step, the action probability of LA is computed based upon the input received from the environment and accordingly the action of LA is performed (lines 36–39). The reward and penalty received by the environment is assumed as proposed in [11–16,31,32]. The ratio of number of times an action is rewarded to the total number of times an action is taken is computed and maximum value of this ratio is taken for maximum benefits (lines 40–43). The complete pseudo code for the proposed algorithm is described as follows:

Efficient Learning Automata based Cell Clustering Algorithm (ELACCA)	
Parameters:	
M :	The number of times an action i is rewarded
K :	Total number of times action i has been taken
A :	Action taken for $i = 1, 2, \dots, l, k$
Output: Cluster head election and connectivity	
Cluster head maintenance	
1.	Set up the initial energy level of the sensor node
2.	Initially select the node having maximum energy as the CH
3.	Assign the automaton to the selected CH
4.	for($S = 1; S \leq n; S++$)
5.	for($level = 1; level \leq \max; level++$)
6.	Calculate the participation degree S^{part} of the nodes in the network using equation (13)
7.	Select the node having maximum participation degree in CH
8.	Calculate the probability of success or failure of cut off in the region as
9.	$P^F = B^i \sum_{j=0}^{i-1} (1 - B^j), P^S = \sum_{i=0}^n (1 - B^i)$
10.	If (LA receives an award) then
11.	The participation degree S^{part} gets incremented as
12.	$S^{part} = S^{part} + p_j(n)$
13.	Else
14.	The participation degree S^{part} gets decremented as
15.	$S^{part} = S^{part} - p_j(n)$
16.	Calculate the cumulative probability of cutoff from equation (5)
17.	Apply the first level cutoff on number of nodes in the selected CH
18.	If ($S^{part} \leq (value(f^{cut}))$) then
19.	Mark this node as the CH for the next level
20.	Else
21.	Perform second level cutoff on number of nodes in the CH according to values of $value(f^{cut})$
22.	End for
Connectivity preservation	

```

23. Calculate the cell radius using equations (7), (8), (9)
24. If (nodes are on the edge of the cell) then
25.     Calculate the distance ( $R_i$ ) between them
26.     Calculate the angle  $\theta$  of the CH from the base position
27.     If ( $(R_i \geq 7.15R_c) \& \& (\theta = 60)$ ) then
28.         CHs on the edges are connected with each other
29.     Else
30.         The CHs on the edges are not connected
31. Else
32.     Find the location of each node using equations (14),(15), (16)
33.     for( $\theta = 10; \theta \leq 90; \theta+ = 10$ )
34.     Find the value of  $R_i$  in terms of  $R_c$ 
35. Find the number of CHs in the network area under investigation

Update of LA Actions probability
36.     If (LA receives an award) then
37.          $A_k(n+1) = \max(A(n), p_j)$ .
38.     Else
39.          $A_k(n+1) = \min(A(n), p_j)$ .
40. Calculate the ratio M/K
41. for each action of LA
42.     Select the maximum value of M/K
43. end for

```

6 Simulation Environment with Results and Discussion

The simulation of the proposed scheme is performed on ns-2 [34] simulator with the parameters as defined below in Table 2. As discussed above automaton are deployed at each of the selected CH which controls all the operations a particular SN performs in the given geographical region. We have compared the performance of the proposed scheme with LEACH [35], EECF [36] and HEED [37]. Various parameters such as cluster size, number of nodes alive, energy consumption are evaluated by varying the number of nodes in the network, with number of rounds for selection of CH are evaluated in the proposed scheme and existing well known schemes. The detailed values of the parameters are listed in the Table 2 below.

6.1 Impact of the Proposed Scheme on Cluster Size

Figure 4 shows the performance of the proposed scheme on cluster size with varying the number of rounds. As shown in the figure below, with an increase in number of rounds for CH selection, the size of the CH is decreased but the proposed scheme contains more nodes than the other existing schemes. It shows the effectiveness of the proposed scheme because with an increase in number of rounds the connectivity of the nodes is higher than the other existing schemes. The reason for the improved performance of the proposed scheme is due to the fact that LA adaptively selects the number of nodes in the CH based upon the participation

Table 2 Simulation parameters

Parameters	Value
Network area	$200 \times 200 \text{ m}^2$
Number of sensor nodes	400
Density of nodes	100 nodes/m^2
Initial energy of the node	0.5 J
Data packet size	512 bytes

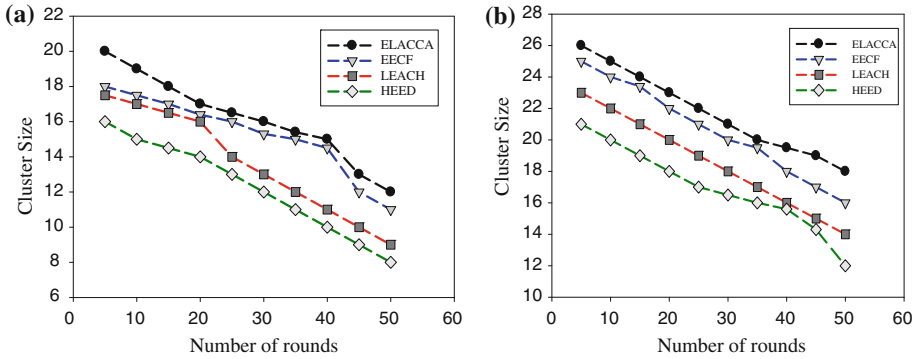


Fig. 4 Proposed scheme with cluster size having a 200 nodes b 400 nodes

ratio of the nodes in the cell. Based upon the participation ratio of the nodes, first level and second level cutoff are applied to the number of nodes in the cell. As illustrated in Fig. 4a, b, with an increase in number of rounds for selection of CHs the size of the cluster is decreased most in the proposed scheme compared to other schemes as cutoff is being applied at different levels based upon the participation ratio of the SNs.

6.2 Impact of the Proposed Scheme on Number of Nodes Alive

Figure 5 shows the impact of the proposed scheme on number of nodes alive after finite number of rounds. Each node in the network has some initial energy but the level of energy decreases as the nodes perform a particular operation and may become zero after finite number of rounds. In such a case the node are considered to be dead and does not perform any action. As shown in the figure, the number of nodes alive in the proposed scheme is higher than the other schemes even high number of rounds. This is primarily due to the reason that in the proposed scheme, LA performs the cutoff operation on the upper bound on the number of nodes in each CH. This will save a lot of energy and increases the life of the nodes in the network. The selection and CH in the other schemes are random and consequently the energy levels of the nodes decrease to considerable amount. Moreover using the defined participation ratio of the nodes in the network the coverage and connectivity is preserved and consequently cutoff at different levels in the network is applied. As LA is deployed on the

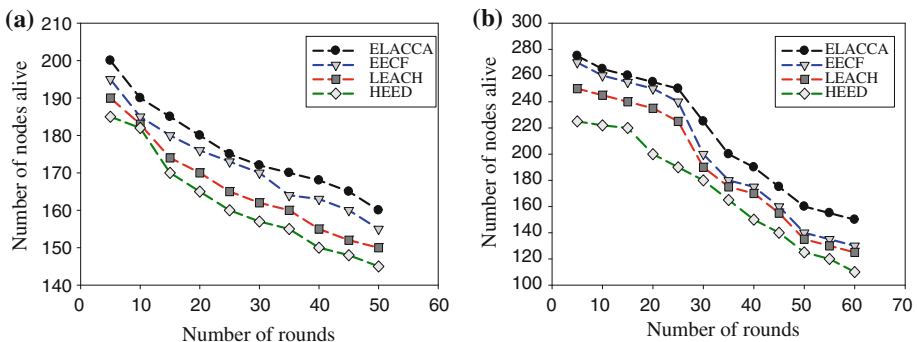


Fig. 5 Proposed scheme with nodes alive. a 200 nodes, b 400 nodes

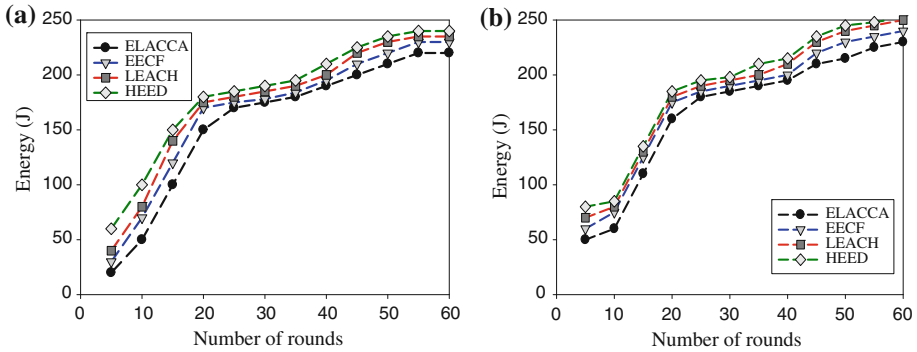


Fig. 6 Proposed scheme on total energy consumed. **a** 200 nodes, **b** 400 nodes

CH so computational complexity is handled by respective LA and not by the individual node. Hence lot of energy of the nodes is saved using the proposed methodology as compared with the other approaches of its category.

6.3 Impact of the Proposed Scheme on Total Energy Consumed

Figure 6 shows the impact of the proposed scheme on total energy consumed with number of rounds. With an increase in number of rounds, the total energy consumption increases in all the schemes for both the cases (a) and (b) of 200 and 400 nodes. But compared to other schemes, the energy consumption is less in the proposed scheme in comparison to other schemes. This is primarily due to the reason that the LA learns from the environment and adapt to the situation quickly and select the best node for maintaining the coverage and connectivity among all the participating nodes. Moreover, LA performs the cutoff operation as defined above with respect to the number of nodes in the CH which further reduces the energy consumption in the proposed scheme compared to other schemes. The advantage of using the cutoff is that only nodes satisfying particular criteria are allowed to participate in CH election which saves lot of energy as compared to other approaches of its category.

6.4 Impact of Cutoff Value on Energy Consumed

Figure 7 shows the impact of cutoff value on energy consumption with varying number of rounds. As seen from the figures, with an increase in the cutoff value in a particular CH, the average energy consumption is also increased. This is primarily due to the fact that with an increase in number of rounds and cutoff, more and more nodes will belong to a particular CH and to maintain all such nodes in the network, lot of energy is consumed which is also reflected in the Fig. 7 also. As shown in the Fig. 7, with an increase in the value of cutoff nodes from 10 to 50, the average energy consumption is increased by 40%. Moreover, some of the nodes are not located in geographical regions where coverage and connectivity is preserved for them, so in the proposed approach estimates about their location with connectivity to the CH is computed and then accordingly major decisions such as selection of CH, coverage and connectivity about the nodes is made.

6.5 Impact of Variation of θ on Number of CHs

Figure 8 shows the impact of variation of values of θ on number of CHs. As we can see with an increase in the value of θ , the number of CHs is decreased to large extent. Moreover, with

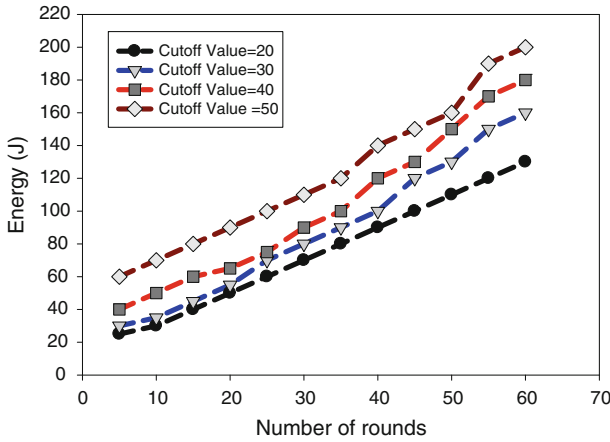


Fig. 7 Variation of energy consumption with cut off value

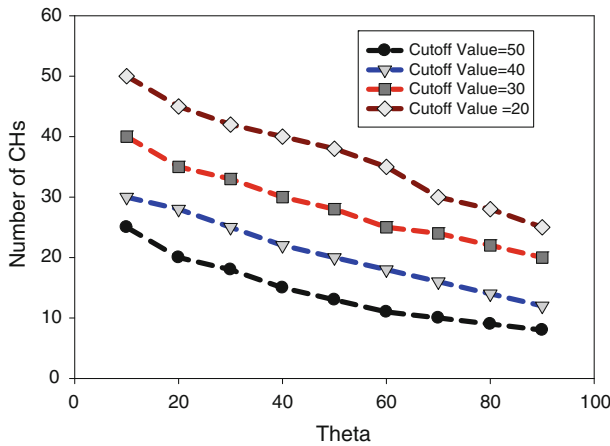


Fig. 8 Variation of θ with number of CHs

an increase in the values of cutoff nodes in the cell, the number of CHs is also decreased. The primary reason for the same is due to the fact that with an increase in the value of more number of nodes are competing to become CH as with smaller values of θ , every node in the cell has chance to become the CH. Moreover, as the value of cutoff is increased more stringent conditions are applied for selection of CH and hence there is a decrease in the number of CHs as can be seen from Fig. 8 below.

7 Conclusions and Future Scope

Clustering with coverage and connectivity in wireless sensor networks (WSNs) are the two key issues which are backbone of many applications in WSNs. But some of the nodes in WSNs may not be able to make contact with the CHs or BS and hence loose communication with them. So an intelligent approach is required in which complexity of the proposed solution

is less. To tackle these issues in WSNs, in this paper we propose a new LA based clustering with coverage and connectivity scheme for WSNs. The area under investigation is divided into various cells having CHs and SNs. A cutoff value is provided which sets the bound on the number of nodes in a particular CH depending upon the participation ratio of SNs. In the proposed scheme, cell shape has been changed to rhombus which is different from the earlier approaches in which it has been assumed as square. The angular displacement from the base of the cell to other points is taken and rotated to give different values of angle. The coverage of maximum number of points from the base of the cell is taken into account. Moreover, Efficient Learning automata based Cell Clustering Algorithm (ELACCA) is proposed. The performance of the proposed scheme is evaluated using various metrics. The results obtained show that the proposed scheme is better than the existing schemes with respect to these metric. In future, we would like to extend this work for secure communication between CHs and SNs.

Acknowledgments This work was supported by the new faculty research program 2013 of Kookmin University in Korea.

References

- Polastre, J., Hill, J., & Culler, D. (2004). Versatile low power media access for wireless sensor networks. In *Proceedings of the 2nd international conference on embedded networked sensor systems, SenSys '04*. ACM, New York, NY, USA, pp. 95–107.
- Alberola, R. D. P., & Pesch, D. (2012). Duty cycle learning algorithm (DCLA) for IEEE 802.15.4 beacon-enabled wireless sensor networks. *Ad Hoc Networks*, *10*, 664–679.
- Kumar, D., Trilok, C., & Patel, A. R. B. (2009). EEHC: Energy efficient heterogeneous clustered scheme for wireless sensor networks. *Computer Communications*, *32*(4), 662–667.
- Yi, S., Heo, J., Cho, Y., & Hong, J. (2007). PEACH: Power efficient and adaptive clustering hierarchy protocol for wireless sensor networks. *Computer Communications*, *30*(14–15), 2842–2852.
- Low, C. P., Fang, C., Ng, J. M., & Ang, Y. H. (2008). Efficient load balanced clustering algorithm for wireless sensor networks. *Computer Communications*, *31*(4), 750–759.
- Fuad, B., & Awan, I. (2011). Adaptive decentralized re-clustering algorithm for wireless sensor networks. *Journal of Computer and System Sciences*, *77*(2), 282–292.
- Zhang, Y., Li, K., Gu, H., & Yang, D. (2012). Adaptive split and merge clustering algorithm for wireless sensor networks. *Procedia Engineering*, *29*, 3547–3551.
- Khalil, E. A., & Attea, B. A. (2011). Energy aware evolutionary algorithm for dynamic clustering of wireless sensor networks. *Swarm and Evolutionary Computation*, *1*(4), 195–203.
- Jin, Y., Jo, J. Y., Wang, L., Kim, Y., & Yang, X. (2008). ECCRA: An energy efficient coverage and connectivity preserving routing algorithm under border effects in wireless sensor networks. *Computer Communications*, *31*, 2398–2407.
- Lin, C., Wu, G., Xia, F., Li, M., Yao, L., & Pei, Z. (2012). Energy efficient ant colony algorithms for data aggregation in wireless sensor networks. *Journal of Computer and System Sciences*, *78*(6), 1686–1702.
- Esnaashari, M., & Meybodi, M. R. (2011). A cellular learning automata based deployment strategy for mobile wireless sensor networks. *Journal of Parallel and Distributed Computing*, *71*(7), 988–1001.
- Esnaashari, M., & Meybodi, M. R. (2010). A learning automata based scheduling solution to the dynamic point coverage problem in wireless sensor networks. *Computer Networks*, *54*(14), 2410–2438.
- Torkestani, A. J., & Meybodi, R. M. (2010). Mobility-based multicast routing algorithm for wireless mobile ad-hoc networks: A learning automata approach. *Computer Communication*, *33*(6), 721–735.
- Torkestani, A. J., & Meybodi, R. M. (2011). Learning automata based algorithms for solving stochastic minimum spanning tree problem. *Applied Soft Computing*, *11*(6), 4064–4077.
- Torkestani, A. J., & Meybodi, R. M. (2010). An intelligent backbone formation algorithm for wireless adhoc networks based upon distributed learning automata. *Computer Networks*, *54*(5), 826–843.
- Chatzichristofis, S. A., Dimitris, M. A., Sirakoulis, G. C., & Boutalis, Y. S. (2010). A novel cellular automata based technique for visual multimedia content encryption. *Optics Communications*, *283*(21), 4250–4260.
- Liu, Z., Zheng, Q., Xue, L., & Guan, X. (2012). A distributed energy-efficient clustering algorithm with improved coverage in wireless sensor networks. *Future Generation Computer Systems*, *28*, 780–790.

18. Aioffi, W. M., Valle, C. A., Mateus, G. R., & Cunha, A. S. D. (2011). Balancing message delivery latency and network lifetime through an integrated model for clustering and routing in wireless sensor networks. *Computer Networks*, 55, 2803–2820.
19. Lai, W. K., Fan, C. S., & Lin, L. Y. (2012). Arranging cluster sizes and transmission ranges for wireless sensor networks. *Information Sciences*, 183, 117–131.
20. Marcelloni, F., & Vecchio, M. (2010). Enabling energy-efficient and lossy-aware data compression in wireless sensor networks by multi-objective evolutionary optimization. *Information Sciences*, 180(10), 1924–1941.
21. Ting, C. K., & Liao, C. C. (2010). A memetic algorithm for extending wireless sensor network lifetime. *Information Sciences*, 180(24), 4818–4833.
22. Bandyopadhyay, S., & Coyle, E. J. (2013). An energy efficient hierarchical clustering algorithm for wireless sensor networks. In *Proceedings of IEEE computer and communications societies (INFOCOM)*, pp. 1713–1723.
23. Wang, Y., Wu, H., Nelavelli, R., & Tzeng, N. F. (2006). Balance based energy-efficient communication protocols for wireless sensor networks. In *Proceedings of IEEE international conference workshops on distributed computing systems*.
24. Alippi, C., Camplani, R., & Roveri, M. (2009). An adaptive LLC-based and hierarchical power-aware routing algorithm. *IEEE Transactions on Instrumentation and Measurement*, 58(9), 3347–3357.
25. Attea, B. A., & Khalil, E. A. (2012). A new evolutionary based routing protocol for clustered heterogeneous wireless sensor networks. *Applied Soft computing*, 12(7), 1950–1957.
26. Yu, J., Qi, Y., Wang, G., & Gu, X. (2012). A cluster-based routing protocol for wireless sensor networks with non uniform node distribution. *International Journal of Electronics and Communications*, 66, 54–61.
27. Sarkar, P., & Saha, A. (2011). Security enhanced communication in wireless sensor networks using Reed–Muller codes and partially balanced incomplete block designs. *Journal of Convergence*, 2(1), 23–30.
28. Pan, R., Xu, G., Fu, B., Dolog, P., Wang, Z., & Leginus, M. (2012). Improving recommendations by the clustering of tag neighbours. *Journal of Convergence*, 3(1), 13–20.
29. Silas, S., Ezra, K., & Rajsingh, E.B. (2012). A novel fault tolerant service selection framework for pervasive computing. *Human-centric Computing and Information Sciences*, 2:5. doi:10.1186/2192-1962-2-5.
30. Dhurandher, S. K., Obaidat, M. S., & Gupta, M. (2012). An acoustic communication based AQUA-GLOMO simulator for underwater networks. *Human-centric Computing and Information Sciences*, 2:3, 2–14.
31. Narendra K. S., & Thathachar, M. A. L. (1980). On the behavior of a learning automaton in a changing environment with application to telephone traffic routing. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC, 10(5), 262–269.
32. Najim, K., & Poznyak, A. S. (1996). Multimodal searching technique based on learning automata with continuous input and changing number of actions. *IEEE Transactions on Systems, Man, and Cybernetics-Part B. Cybernetics*, 26(4), 666–673.
33. Luo, H., Luo, J., Liu, Y., & Das, S. K. (2009). Adaptive data fusion for energy efficient routing in wireless sensor networks. *IEEE Transactions on computers*, 24(5), 345–359.
34. The Network Simulator NS-2. <http://www.isi.edu/nsnam/ns/>.
35. Chandrakasan, A. P., Smith, A. C., & Heinzelman, W. B. (2004). An application specific protocol architecture for wireless micro sensor networks. *IEEE Transaction on Wireless Communications*, 1(4), 660–669.
36. Chamam, A., & Pierre, S. (2010). A distributed energy-efficient clustering protocol for wireless sensor networks. *Computers & Electrical Engineering*, 36(2), 303–312.
37. Younis, O., & Fahmy, S. (2004). HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *IEEE Transactions on Mobile Computing*, 3(4), 366–379.

Author Biographies



Neeraj Kumar is working as Assistant Professor in Department of Computer Science and Engineering, Thapar University, Patiala, Punjab (India). He received his Ph.D. in CSE from Shri Mata Vaishno Devi University, Katra (India) and PDF from UK. He has more than 40 research publications in peer reviewed journals and conferences including IEEE, Elsevier, and Springer. His research is focused on mobile computing, parallel/distributed computing, multiagent systems, service oriented computing, routing and security issues in wireless adhoc, Sensor and Mesh networks. He is leading the Mobile Computing and Distributed System Research Group. Prior to joining Thapar University, Patiala, (Punjab), he has worked with HEC Jagadhri (Haryana) and MMEC Mullana, Ambala (Haryana), and SMVDU, Katra (J&K). He has delivered invited talks and lectures in various IEEE international conferences in India and abroad. He has organized various special sessions in international conferences in his area of expertise in India and abroad. He is TPC of various IEEE sponsored conferences in India and abroad. He is reviewer/editorial board of various international journals

e.g. IEEE TVT, IEEE Transaction on Wireless Communications, Journal of Network and Computer Applications (Elsevier), Wireless Personal Communications (Springer), Future Generation Computer Systems (Elsevier), Computers and Security (Elsevier), Mathematical and Computer Modeling (Elsevier), Information Fusion (Elsevier), Eurasip Journal of Wireless Communication and Networking (Hindawi), International Journal of Network Security (IJNS), Journal of Emerging Trends in Web Intelligence, Journal of Advances in Information Technology, IJCA and many more. He has Edited/Editing special issue of 6 international journals from Springer, Inderscience, and Academy publishers. He is senior member of ACEEE and IACSIT.



Jongsung Kim received his Bachelor and Master degrees in Mathematics from Korea university, Korea in 2000 and 2002, respectively. He received double Doctoral degrees completed in November 2006 and February 2007 at the ESAT/COSIC group Katholieke Universiteit Leuven and at Engineering in Information Security of Korea University, respectively. He had been a Research Professor of Center for Information Security Technologies (CIST) at Korea University, Korea, from March 2007 till August 2009, and an assistant professor of department of e-business at Kyungnam University, Korea, from September 2009 till February 2013. Dr. Kim has been an assistant professor of department of mathematics at Kookmin University, Korea, since March 2013. Dr. Kim has published about 60 research papers in international journals and conferences. He has been serving as chairs, program committee, or organizing committee chair for many international conferences and workshops. He is editorial board member of International Journal of Information Technology, Communications and Convergence (IJITCC), International Journal of Communication Networks and Distributed Systems (IJCNDS), InderScience and Journal of Convergence (JoC), FTRA Publishing, and Human-centric Computing and Information Sciences (HCIS), Springer. In addition, he has been serving as a Guest Editor for international journals by some publishers: Springer, Elsevier, John Wiley, Oxford Univ. press, Inderscience. His research interests include security issues, cryptography, ubiquitous computing systems and digital forensics.

tems (IJCNDS), InderScience and Journal of Convergence (JoC), FTRA Publishing, and Human-centric Computing and Information Sciences (HCIS), Springer. In addition, he has been serving as a Guest Editor for international journals by some publishers: Springer, Elsevier, John Wiley, Oxford Univ. press, Inderscience. His research interests include security issues, cryptography, ubiquitous computing systems and digital forensics.