Improving lifetime of wireless sensor networks based on nodes' distribution using Gaussian mixture model in multi-mobile sink approach

Houriya Hojjatinia[1] · Mohsen Jahanshahi[2] · Saeedreza Shehnepoor[3]

Abstract

Saving energy in Wireless Sensor Networks (WSNs), is critical in different applications, such as environment monitoring, keeping human awareness and etc. Many studies have investigated energy consumption and improved the WSN lifetime longevity by reducing the energy consumption. Still, proposed approaches overlook the nodes' distribution role in energy model and routing protocol, which is a key factor in a WSN. In this work, we propose a novel approach; namely *GDECA*; which assumes nodes' distributions are mixtures of Gaussian distribution, as an assumption applied in real world. So *GDECA* rely on a distribution estimation borrowed from Machine Learning (ML) to fit the Gaussian Mixture Model (GMM) to the nodes and calculate the parameters for these distributions. Next, the estimated parameters are employed in Cluster Head *CH* selection policy. Besides, sinks routing is determined based on nodes distribution. Results showed the improvement close to 40–50% in energy consumption. As another outcome, GDECA keeps all the nodes active until end of the simulation. Observations also demonstrate that sinks path calculation using this approach is optimum, and randomly changing number of sinks increases energy consumption.

Keywords Wireless sensor network · Gaussian mixture model · Cluster head selection · Energy model · Energy consumption

1 Introduction

Wireless Sensor Networks (WSNs) consist of wireless sensors with specific transmission range to send and receive information and to perform simple mathematical operations, locally [1]. Such sensors are occupied in different open field applications; such as environment monitoring, events management, etc. for context awareness [2].

To address recharging limitations of wireless sensors, several algorithms and protocols are proposed to save sensors energy, and finally to increase lifetime of WSNs. On the other hand, such massive number of the algorithms and protocols is unable to solve all the challenges in terms of energy, computational power, memory usage and etc.

Two types of strategies are proposed to improve the energy consumption in WSN; first solutions consider deployment of the nodes in WSN as a solution to solve the energy consumption. Such studies investigate the optimum solutions for nodes' deployment which can be helpful to reduce the energy consumption [3–5]. The proposed approach for nodes deployment benefit from suitable nodes' distribution over the network to reduce energy consumption. Apart from these approaches, second type of techniques employ routing protocol techniques used to save energy for each node. In contrast to first category, this type of techniques assumes that node's distribution is arbitrary and focuses on the techniques to find best routing protocol for the sinks to increase network lifetime. In current study, we focus on second type to address the energy consumption problem in a WSN.

✉ Mohsen Jahanshahi
  mjahanshahi@iauctb.ac.ir

  Houriya Hojjatinia
  houriya.hojjatinia@srbiau.ac.ir

  Saeedreza Shehnepoor
  saeedreza.shehnepoor@research.uwa.edu.au

[1] Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

[2] Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran

[3] Department of Electronic, Electrical and Computer Engineering, University of Western Australia, Perth, Australia

One-hop and multi-hop protocols suffer from different drawbacks such as "hotspot" problem; which refers to overhead on central sink that is responsible to collect data from other nodes. This overhead leads to delay in sending packets to receivers. Data collection approaches such as "Hierarchical Protocols" are proposed to save sensors' energy and distribute energy waste among all the sensors in networks, uniformly and also to increase the scalability of the network [6,7]. We can refer to Mobile based Energy efficient Clustering Algorithm (MECA) and Energy efficient Multi-sink Clustering Algorithm (EMCA) as solutions to address this type of problems [8].

Although many studies considered hierarchical and clustering algorithms, but such approaches still overlook different limitations. Initially, selecting only one node as CH in algorithm leads to overhead on CH and this eventually leads to nodes' energy drainage. Initial studies proposed different approaches to select CH, while recent studies randomly selected CH with no pre-determined strategy. One important defect of previous algorithms is that they not only ignore distribution of nodes in network in first place, but also ignore distribution of nodes in each cluster. In better words, recent approaches suffer from serious limitations. First, the latest techniques overlook the key role of nodes' distribution in determining sinks' movement. Estimation of nodes' distribution simply helps the sinks to move regarding the nodes' distribution. Furthermore, recent works are limited to narrow scope of utilizing criteria based on nodes' energy to determine CH. Considering nodes energy for CH selection partially addresses the "hotspot" problem, but still we require the spatial information of nodes to better select the CH. As an example, selecting the CH in most dense regions simply reduce the energy consumption, since the average distance between nodes and CH is decreased. Addressing both limitations result in WSN lifetime longevity.

First we require to cluster nodes and select CH based on a possible distribution of nodes. In a same way, sinks' path is also determined. In addition, we devise a new packet transmission scheduling for sinks.

For distribution estimation, we assume nodes distribution as Gaussian Mixtures and hence we employ Gaussian Mixture Model (GMM) from Machine Learning (ML) area [9] to calculate each cluster using Mean and Covariance as distributions' parameters. This model uses Expectation-Maximization (EM) to calculate each cluster, while maximizing node distribution likelihood in each iteration. This algorithm maximizes Gaussian posterior distribution until convergence is satisfied. Afterwards the nodes close to mean value are selected as CH. Hence, we can have CH closer to nodes density and therefore we can improve their energy consumption, considering reduction in average distance between nodes and their CH.

On the other hand we also determine the sink path based on the nodes distribution and hence we can select a path with least energy consumption for sending packets. So we can summarize our contributions as follows:

– For the first time, we employ a clustering method borrowed from Machine Learning (ML) [9] (GMM) to estimate the nodes' distribution using latent variables (See Sect. 3.2). To this end, the model takes the number of mixtures as input and estimates the nodes in each cluster and statistical information as the information required for routing protocol.
– In contrast with previous works, we use the statistical information from estimation to calculate the routing path for each sink. We utilize the intersection between each two clusters to determine the intended routing path. The results shows that the proposed approach improved the energy consumption and lifetime of the network (See Sects. 5.1.2, 5.1.3, 5.1.4).
– We propose a novel CH selection algorithm, which utilizes the statistical information acquired from GMM to reduce the distance between nodes and CH. We prove that such an approach is optimum (See Eq. (17) in energy consumption. We also reflects the effectiveness of proposed model in reducing the average distance in Sect. 5.1.1.

2 Related works

In the last decade, a great deal of research studies focus on proposing new algorithms to reduce the energy consumption. Some focused on nodes' deployment to increase the lifetime of WSN, and others focused on developing the routing protocols, able to reduce the transmission energy required to send data to sink and CH. In this section, we first introduce the studies on nodes deployment studies and then move to focus of our study; routing protocols.

2.1 Sensor deployment using Gaussian distribution

Different studies considered the deployment of a WSN using Gaussian distribution. A survey in signal level of techniques used in WSN is studied by Yetgin et al. present different algorithms proposed to increase Network Lifetime (NL) of WSN. Different guidelines are provided to show the possible directions of future works.

Halder et al. propose a deployment based on customized Gaussian distribution to address the standard Gaussian distribution problems in balancing the energy consumption. The results showed an outstanding improvements in terms of lifetime of WSN and energy consumption.

An analytical framework is proposed by Wang et al. [4] to consider 2D Gaussian to improve the deployment of nodes

and introduce new parameters on Gaussian model to improve the energy consumption in WSNs. Proposed approach effectively improve the lifetime of network.

Boukerche et al. [10] propose an approach in corona based WSN for both uniform and non-uniform deployment of nodes. In proposed approach nodes calculate the probability of delivering the data to other nodes in different situations. The results show that performance of the deployment approach is significantly improved.

2.2 Data collection approaches

A review of data collection approaches is given in Fig. 1.

As displayed in Fig. 1, given the significance of sink mobility, the approaches are divided into two groups, at the first level; fixed sink and mobile sink approaches. The former is categorized based on the number of sinks responsible for collecting the information from the nodes; multi-sink and single sink. The latter approach is divided into three categories according to the number of intermediate sinks contributed to the routing task; one-hop, multi-hop, and hierarchical routing protocol. Among three mobile sink approaches, the hierarchical protocol is divided to multi-sink and single sink approaches. Each category includes three types of management strategy employed by sinks to direct packets to Access Points (APs); controlled, random, and deterministic. In following we explain different approaches and provide various examples for each category. The fixed sink category is divided up into two sub-categories: single sink; where data are delivered to a single sink, directly [11], and multi sink approaches; where data are received by target in multiple hops with intermediate nodes [12]. Fixed, single sink is useful when covered wireless network is small, while multi-sink type is suitable when multiple sinks are available to gather data. As a result this category is applicable to vast areas. In this approach network is divided to areas with different or same sizes and every sink receives data from nodes or CH [13].

Second category refers to approaches which assume that sinks are mobile. Different studies demonstrated that mobile sinks improves the efficiency in terms of different metrics [14]. Mobile sinks approaches proposed to increase the network lifetime [15], cover the networks [16], increase the nodes' communications [17], and decrease the packet loss [18].

Generally, there are three approaches for collecting data, for mobile sinks. In first approach, sinks collect data from nodes, directly. Obviously such approaches are suitable for small networks [19].

Second category consists of approaches which are called "Multi-hop" routing protocols. Intermediate nodes play important roles in transmitting data to destination in this category, but they make no changes in sent packets data. This approaches are known for their contribution in large networks [20].

Third approach is known as "Hierarchical Approach"; suitable for networks with tens of thousands nodes. In this approach, networks are divided to different layers and each layer is leveled with different types of networks [21,22]. In each layer, collected data is transmitted to a CH and each CH sends data to the upper layer. Consequently, approaches in this category are using the multi-hop routing. Different criteria is used for evaluating approaches in this area. Some approaches, select CH based on node send/receive data range. Remaining studies select CH based on residual energy of nodes in clusters. In a hierarchical approach, single or multiple sinks are selected based on network scale. Single sink approach is not suitable for large networks or event based data. On the other hand, multiple sinks approach decrease delay in data collection [23] of large networks, increase in data throughput [24], and energy saving for nodes [25].

Routing protocols generally consider energy consumption distribution among nodes, uniformly. To address this problem, algorithms require to solve hotspot problem and hence, increase network lifetime. Several approaches used clustering as a solution for energy consumption [8,26,27]. Clustering approach will reduces intra-cluster communication range and as a result, energy consumption reduces.

The multi-hop communication, consume lower energy than one-hop communication in large networks [28]. So, energy consumption is distributed among nodes and this prevents nodes' death in short time (hotspot problem) [27,29].

Different protocols are introduces to find the optimum path to CH to save the energy such as Multiple Mobile Sink-based Routing (MMSR), Energy-efficient Multi-sink Clustering Algorithm (EMCA), Mobile-sink based Energy-efficient Clustering Algorithm (MECA), Tour-Planning Algorithm (TPA), Energy efficient Distance-Aware Routing Algorithm (EDARA), and etc.

MMSR is proposed to reduce difficulties related to hotspot problem and increase network lifetime. In this algorithm WSN ares is divided to two sections, internal circle, and external circle. The external circle is divided to multiple areas. A mobile sink moves along circle's diameter, alongside two other sinks moving on circle's arc.

In EDARA, multiple mobile sinks are engaged to collect data. In EDARA, sinks are fixed in their initial point. Then, first sink starts moving from initial point, and other sinks start moving, consequently. Sinks will not collect any data until they get to the next station. Starting movement again, sinks inform the sinks about their location. Sensor nodes send data to destination, using one-hop, or multi-hop approach. Information overhead is one drawback for this algorithm, since in every station, each sink broadcasts the location data to each node using a hand-shaking algorithm.

Fig. 1 Structural chart of the different approaches which were employed in previous data collection methods (Note that Deter. stands for Deterministic)
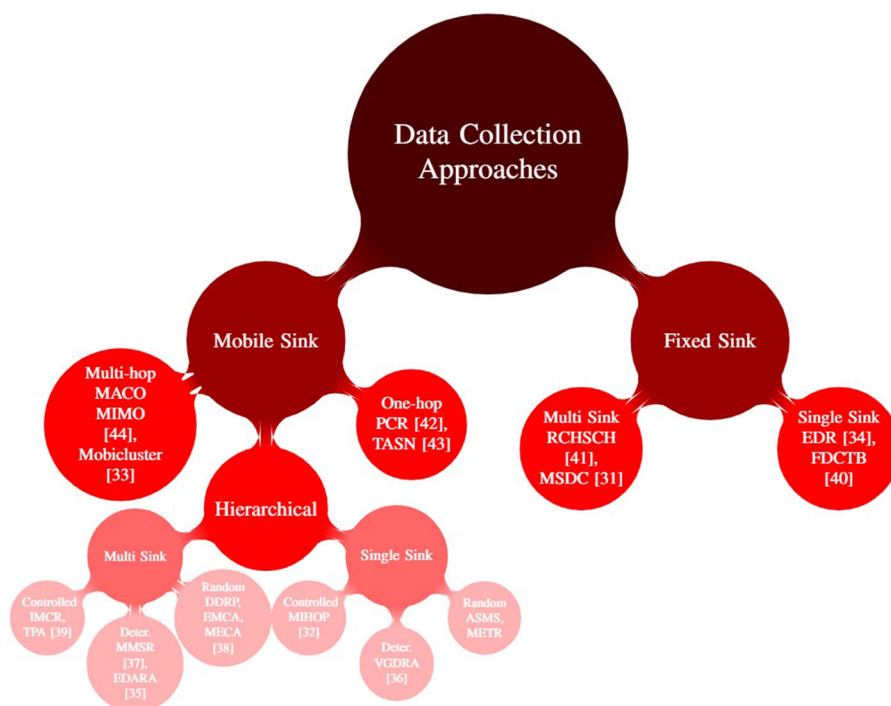
WSN is divided to sub-networks in TPA, and a mobile sink collects data from sensors, directly. TPA preselect several nodes as intermediate nodes to transmit data to CH. Next, the algorithm estimates shortest path to destination, via multiple nodes. TPA is applicable to small networks, due to limited transmission range.

As a fixed multiple sinks approach, EMCA uses random distribution for sinks across network area. In this approach, network is divided to multiple clusters, and the node with most residual energy is selected as CH. Nodes send information to CH, in one-hop or multi-hop approach and CH choose optimum sink for information transmission. Nevertheless, hotspot problem is still a drawback for EMCA, since sinks are fixed in their locations.

In contrast to EMCA, multiple mobile sinks are engaged in MECA [8]. Sinks velocity is predetermined and fixed during algorithm execution, and move along a circle's perimeter. Sink broadcasts the location only once starting the algorithm, and nodes will calculate sinks' next location based on their initial location and velocity. In MECA, network is divided to multiple equal clusters, and CH is selected based on residual energy for each node. MECA adopts one-hop or multi-hop approach to send data to targeted node. In multi-hop approach, achieving energy saving is accessible, since nodes are able to send data to CH or sink, directly or in multiple steps. Amini et al. [30] proposed an improved version of MECA, where the area of distributed nodes was considered as hexagon. To improve the effectiveness of the model, sinks move inside the hexagon and stoppage locations are deter-

mined inside the hexagonal area. Using such approach, the performance of the proposed model has improved in terms of network lifetime and energy consumption.

In conclusion, previous approaches overlook nodes' distribution in CH selection, and sink movement path. Considering nodes distribution, hotspot problem will be almost solved, since CH is selected based on the vicinity to dense area. In addition, balance on energy consumption among nodes is other achievement, since CH are selected periodically. Moreover, propagation delay is reduced and throughput is increased, since nodes send information to CH in dense areas. Finally, isolated nodes can be also available in algorithm and hence, network performance will be improved. The current proposed approach; GDECA is considered as a mobile sink strategy which employs the hierarchical routing protocol in order to manage multiple sinks with controlled algorithm.

GDECA leverages the effectiveness of nodes' deployment introduced in Sect. 2.1 and adopts the nodes' distribution in determining the routing protocol. Hence, we utilize both categories in order to improve the lifetime of WSN.

3 Preliminaries

In this section, we explain assumption, definitions, basic concepts, and base model for our proposed model. First, assumptions for our model is introduced. Then we discuss base model used as energy consumption model for this work. Finally, Gaussian Mixture Model (GMM) is discussed as a model used to cluster the nodes distributed over WSN.

3.1 Model assumptions

Different assumptions is presented in current study. Some of assumptions ensemble different aspects of network in real world, and others are assumed to help problem modelings and simplification. To investigate the impression of each assumption, first we require to identify the quantifiable assumptions. Hence, we define two types of assumptions; hard assumptions and soft assumptions. In following we define each type of assumptions.

Hard assumptions: such assumptions limit the physical and computational capacity of a WSN. Hard assumptions are summarized as below:

– The nodes in network are all fixed and they have no movement during the simulation.
– Concerning computational capacity of nodes, we assume that every random node selected can solve problem solved in polynomial time.
– In addition, nodes have sufficient capacity to transmit data to other nodes, with no disruption.

Soft assumptions: such assumptions refers to routing algorithms employed in a WSN:

– In this work nodes are distributed randomly with no pre-assumption on distribution. To model the distributions we use GMM. This assumption is regarding real world, e.g. cities, where nodes are more dense in some places and they are faded away, while distance from denser parts increases. We can consider every network as a mixture of such model.
– The sinks move in both directions, with a fixed velocity, and their path can be determined based on nodes' distribution.
– The sinks movement starts when their starting and ending point is determined. Stoppage stations is 5 for each sink. It will be done by node that calculates GMM for nodes. After calculation, start and end point for each sink is broadcasted by the node. Remaining nodes and sinks will receive this message and calculate the stoppage stations. Using available data and also the velocity for each sink, each node is aware of sinks' stoppage time, enabling the nodes to send packets in real time.

Note that the soft assumptions are quantifiable and to investigate their contribution in model's performance, we devised three sets of scenarios. These scenarios are introduced and defined in Sect. 5.3. The above assumptions are among base assumption considering for network initialization and simulation. Second assumption is base assumption for novelty in this work and there are lots of clues to support it.

3.2 Gaussian mixture model

As we discussed, nodes in network are distributed based on mixtures of different distribution; most likely Gaussian model. In this work, nodes are distributed with X as their location, M as number of mixture models, θ_m as parameters for each distribution where m stands for mth mixture model. In GMM estimation, an Expectation Maximization (EM) approach is used to predict parameters of each model (each Gaussian model is described using two parameters; mean and variance) [9]. So, we require to maximize $p(X|\theta_m)$, and hence we can obtain probability of a node to see whether it belongs to a model to a Gaussian model or not. For this purpose, a hidden variable, z_{mn}, equals 1, if node x_n belongs to mth mixture model, and otherwise it takes 0. In better word, z_{mn} determines which node belongs to which mixture. First, we rewrite $p(X|\theta)$ as follow:

$$p(X|\theta) = \sum_Z p(X, Z|\theta) \tag{1}$$

We use log for both sides and multiply and divide the right side by $p(Z|X)$:

$$\log p(X|\theta) = \log \sum_Z p(Z|X) \frac{p(X, Z|\theta)}{p(Z|X)} \tag{2}$$

Since the probability function is convex, we have:

$$\log E(x) \geq E(\log(x)) \tag{3}$$

Considering 3 we can write Eq. 2 as:

$$\log \sum_Z p(Z|X) \frac{p(X, Z|\theta)}{p(Z|X)} \geq \sum_Z p(Z|X) \log \frac{p(X, Z|\theta)}{p(Z|X)} \tag{4}$$

Both Z and X are independent variables, so joint probability for them is defined as multiplication of their probability. As a result, right side of (2) is can be written as below:

$$\sum_Z p(Z|X) \frac{p(X, Z|\theta)}{p(Z|X)} = \sum_Z p(Z|X) \frac{p(X|\theta) p(Z|\theta)}{p(X|\theta)} \tag{5}$$

In Eq. 5, term $p(Z|\theta)$ show the probability of z_{mn} equal to 1 for each x_n. This can be considered as $p(m)$. So, we can rewrite right-hand-side (rhs) of Eq. 5 as follow:

$$\sum_{m,n} p(m|x_n) \log \frac{p(x_n|\theta_m) p(m)}{p(m|x_n)}$$
$$= \sum_{m,n} p(m|x_n) \log p(x_n|\theta_m) p(m) \tag{6}$$
$$- \sum_{m,n} p(m|x_n) \log p(m|x_n)$$

EM algorithm tries to maximize the likelihood of two variables Z and X. This algorithm consists of two main steps (as its name demonstrates):

– Expectation: calculating the probability of each node belonging to each cluster.
– Maximization: calculating the probability of x_{mn} equal to 1 for x_n; $p(m)$.

The output of first step is used for second step. Then we calculate $p(m)$ which helps to calculate probability of belonging each node to each cluster and the algorithm is repeated until parameters converge.

To calculate initial values for first step, Eq. 6 is optimized with respect ot $p(m|x_n)$:

$$\frac{\partial LL}{\partial p(m|x_n)} = \log p(m|x_n) - \log p(x_n|\theta_m) p(m) - 1 \qquad (7)$$

So if Eq. 7 equal to zero, then we have:

$$p(m|x_n) = \frac{p(x_n|\theta_m) p(m)}{\sum_{m'} p(x_n|\theta_{m'}) p(m')} \qquad (8)$$

In maximization step, hidden variables' probability is maximized regarding θ_m. Obviously, first term of rhs has a partial difference non-equal to zero. In addition, $p(x_n|\theta_m)$ is considered as multivariate Gaussian distribution. So, the term is substituted with multivariate Gaussian distribution, and differentiate regarding mean, and variance:

$$\mu_m = \frac{\sum_n p(m|x_n) x_n}{\sum_n p(m|x_n)} \qquad (9)$$

$$\Sigma_m = \frac{\sum_n p(m|x_n)(x_n - \mu_m)^T (x_n - \mu_m)}{\sum_n p(m|x_n)} \qquad (10)$$

Then value for $p(m)$ can be calculated using following equation:

$$p(m) = \frac{1}{N} \sum_n p(m|x_n) \qquad (11)$$

Hence, the expectation calculation is completed, and these two steps are repeated until parameters converge to fix values.

4 GDECA; GMM based distribution estimation clustering algorithm

In this section, the proposed model is discussed in two subsections Base model, and *GDECA*. In first subsection base energy model is discussed based on the algorithm proposed in [8] and then we explain the improved energy model.

4.1 Base energy model

In this work, we consider the network as a graph $G(N,E)$, where N indicates nodes in our network, and E is the link between each two nodes (i,j). Every node in network has a communication range (d_0), which enables the node to send data with lower energy and energy consumption increases beyond the range, exponentially. So, we can compute energy consumed to send l-bit from node i to node j located with distance d from each other using following equation:

$$E_{Tx}(l, d) = \begin{cases} lE_{elec} + l\epsilon_{fs}d^2, d < d_0 \\ lE_{elec} + l\epsilon_{mp}d^4, d \geq d_0 \end{cases} \qquad (12)$$

In Eq. 12, the model considers ϵ_{fs} if free space model parameter is lower than node's communication range. If their distance is higher than their communication range, ϵ_{mp} is used as model parameter. For receiving, we use simpler equation as below:

$$E_{Rx} = lE_{elec} \qquad (13)$$

Equation 12 is employed to calculate energy required to transmit packets from node S_i to a node as Cluster Head of ith cluster:

$$E_1(S_i, CH_{S_i}) = \begin{cases} lE_{elec} + l\epsilon_{fs}d(S_i, CH_{S_i})^2, d < d_0 \\ lE_{elec} + l\epsilon_{mp}d(S_i, CH_{S_i})^4, d \geq d_0 \end{cases} \qquad (14)$$

As Eq. 14 indicates, it is costly for a node to send its packets to a CH with high distance. So multi-hop routing is used to send its packets through an underlying node named as S_j to CH_{S_i}:

$$E_2(S_i, S_j, CH_{S_i}) = E_{Tx}(l, d(S_i, S_j)) \\ + E_{Rx}(l) + E_{Tx}(l, d(S_j, CH_{S_i})) \qquad (15)$$

Each node selects the underlying node with minimum energy consumption in multi-hop routing:

$$E_2(S_i, S_j, CH_{S_i}) = \min(E_2(S_i, S_j, CH_{S_i})) \qquad (16)$$

We also calculate energy consumed to send packets to sink directly using energy model in free space and multi path routing and name it as $E_3(S_i, BS_K)$. This is because some nodes can send packets, with less energy when send send their packets to sink directly,rather than sending the packets using multi hop routing or sending packets to CH.

After Calculating E_1, E_2 and E_3, node decide to send packets by approach with minimum consumed energy.
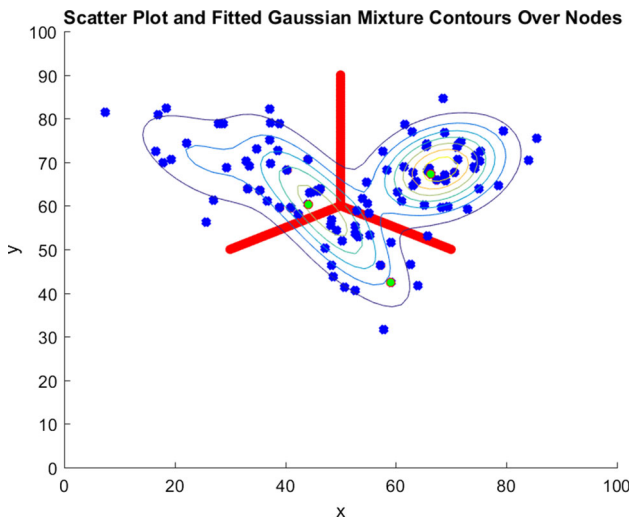
Fig. 2 Network configuration and sink paths (nodes are determined by blue, sinks path by red and CHs by green)

4.2 Proposed approach

In proposed approach, the nodes first broadcast the locations to other nodes, so the rest of the nodes will be aware of their locations. After broadcasting, one of nodes starts to use GMM introduced in Sect. 3.2 to calculate clusters' mean and variance. All nodes are capable of performing this operation, but selected node is random. This node require two types of information:

– Nodes' location.
– Number of mixture models.

First information is acquired by broadcasting step. Second one is predetermined in network and can be changed by configuration setting related to network. Using this information, means and variances are calculated and broadcast by node to sinks and other nodes.

After that sinks will calculate their paths based on intersections between each two GMMs and sinks start to move between intersections based as start and end point. Figure 2 shows their path, start and end point.

This path determination helps CHs to send their packet with less energy, as we explained before. each mixture model is showed as circles in Fig. 2 where higher temperature shows the more dense regions of distribution and lower temperature shows the less dense regions in distribution.

Nodes calculate sink stoppage stations given the start and end point coordinates provided by sinks, the velocity and number of fixed stoppage stations.

As we mentioned, selecting closest node to mean as CH, is less energy costly than selecting the CH after former CH death. Equation 14 shows a direct relation between energy

consumption and distance to CH. A simple proof shows that node located closest to cluster mean is the optimum selection of first CH. For optimizing energy consumption, we need to differentiate energy consumption regarding to distance, since distance is the main variable to determine the energy consumption. So we do summation over all nodes' energies for a cluster and differentiate:

$$E_{overall} = \sum_i l E_{elec} + l \epsilon_{fs} d^2(i, CH)$$

$$\frac{\partial E_{overall}}{\partial d} = \sum_i 2d(i, CH) = 0$$

$$\sum_i 2d(x_i - x_{CH}) = 0 \qquad (17)$$

$$\sum_i x_{CH} = \sum_i x_i$$

$$x_{CH} = \frac{1}{N} \sum_i x_i$$

Where x_i and x_{CH} display nodes' location and CH location, respectively, and N indicates the number of nodes in each cluster.

As Eq. 17 shows, nodes' mean location is optimum location to select a node as initial CH. So in every stoppage closest node to mean is selected as CH. In next station, next closest node to mean is selected as CH and to prevent hotspot problem. The algorithm repeats on the same basis until all nodes are selected once as CH, and the same loop repeats again. If a node is out of energy and off, next closest node is selected as CH.

5 Experimental evaluation

This section compares proposed approach with last state-of-the art solution presented in [8,30]. Next, observations are discussed and analyzed based on different criteria to show proposed approach performance in different circumstances.

5.1 Comparative results

In this section, *GDECA* is compared against the most recent state-of-the art approaches with resepect to different metrics. In next sections, first we discuss the average distance from CH rate, residual energy, total consumed energy regarding different number of nodes.

Network configuration is explained in Table 1. In different sections, different metrics are changed in order to evaluate performance and analyze the observations.

Table 1 Network configuration in proposed approach

Symbol	Name	Value
R_h	Length of the square sideway	45 m
N	Number of nodes	100
M	Number of sinks in the hexagon	6
R_s	Communication range of the sink	20 m
R_n	Communication range of the node	1.5 m
E_0	Initial energy	0.5 Joul
l	packet length	64 bits
ϵ_{fs}	Free-space channel parameter	0.0000015 $pJ/bit/m^2$
ϵ_{mp}	Multi-path channel parameter	0.0000013 $pJ/bit/m^4$
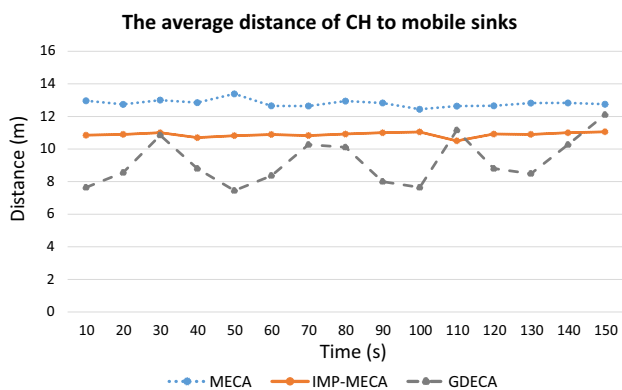
Fig. 3 Energy consumption in [8,30] versus *GDECA* for 100, 200, and 300 nodes
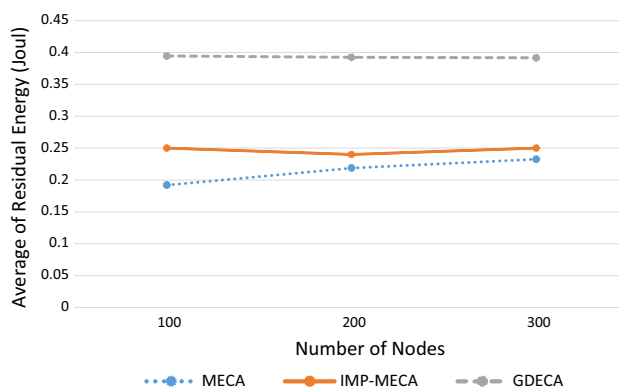
Fig. 4 Average Residual Energy for 1000 seconds simulation in [8,30] versus *GDECA* for 100, 200, and 300 nodes

5.1.1 Average distance rate

As Fig. 3 indicates, *GDECA* outperform [8,30] in at different points of simulation in energy consumption point of view. When simulation starts, average distance is keeps the minimum value, since selected CH is the optimum choice (As in Eq. 17), but with time progress, average distance increases too. One possible explanation is that as we explained in Sect. 4.2, in each stoppage station, the distance between CH and cluster mean increases. As a result, the consumed energy to send packets from nodes to CH increases. Such situation leads to more energy consumption. With all nodes selected as CH in one iteration, close node to mean is selected as CH and again and average distance is reduced (As shown in $t = 50$ in Fig. 3). The interesting point is that after every 50 seconds period, the average distance is repeating the same pattern. This pattern repeats and repeats until all nodes are exhausted and dead. Obviously, in each period, the average distance is increasing, since the distance between new CH and mean node is increasing. This explains why average distance has a periodic shape over time.

Although the average distance of nodes to sinks is changing frequently, the average distance is reduced, significantly.

It is worth mentioning that average distance, is 30%, and 15% reduced compared with [8,30], respectively.

5.1.2 Average of residual energy

One of most important factors contributed to lifetime of a WSN is residual energy, indicating how long the network survives. As it is shown in Fig. 4, average residual energy of *GDECA* for the period length of 1000s, is closely twice as average residual energy for *MECA*. Different reasons are contributed to such performance. First, in *MECA* the node selected for CH has highest residual energy, and no location information is used to calculate CH. Obviously every selected node would consume different amount of energy depending on their distance to other nodes. On the other hand, in *GDECA* every selected node is selected based on its distance to other nodes to optimize energy consumption. Besides, selection method for CH selection, guarantees energy consumption drop. This helps proposed approach to be effective twice as *MECA* in energy consumption.

5.1.3 Number of active nodes

As it is shown in Fig. 5, all nodes survives through the simulation time, while in *MECA*, several nodes are dead. This
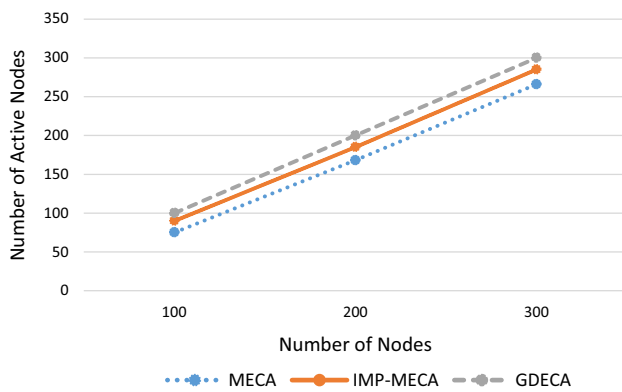
Fig. 5 Number of Active nodes for 1000 seconds simulation in [8,30] versus *GDECA* for 100, 200, and 300 nodes

Fig. 6 Total energy consumption for 1000 seconds simulation in [8,30] versus *GDECA* for 100, 200, and 300 nodes

happens due to uniform pattern in energy consumption. In *MECA*, nodes are selected only regarding to the residual energy. Such selection is costly if a node is distant from other nodes. On the other hand, *GDECA* has deployed a mechanism, helping nodes to have same energy consumption distribution, and this helps network lifetime increament. Finally, the proposed model has also helped the sinks to determine their path based on node's distribution, not based on a pre-determined assumption. Such controlled mechanism helps nodes to be flexible in sending the data to their desired sink. In addition, WSN can be extended with new number of nodes, without affecting the performance, since the sinks routing can be adapted based on new location.

5.1.4 Total energy consumption

Total energy consumption is also regarded as one of the most important factor to show superiority of the proposed approach. Figure 6 indicates that total energy consumption of nodes increases given more nodes in a network. Obviously, total energy is completely dependent on number of nodes, but total energy in different number of nodes, has advantage over *MECA*, and *IMP-MECA* close to 200%, and 100%, respectively.

5.2 Observation analysis

In this section, effects of our proposed approach on WSN is investigated respecting three different criteria. First, we investigate effects of different number of mixture models on performance of *GDECA*. We also analyze impression of sinks paths determination based on number of distributions on energy consumption. In addition, effects of sink number on energy consumption, is also studied to observe the number of nodes' contribution to the energy consumption. Finally, stoppage time impression on energy consumption, and remained active nodes is discussed.
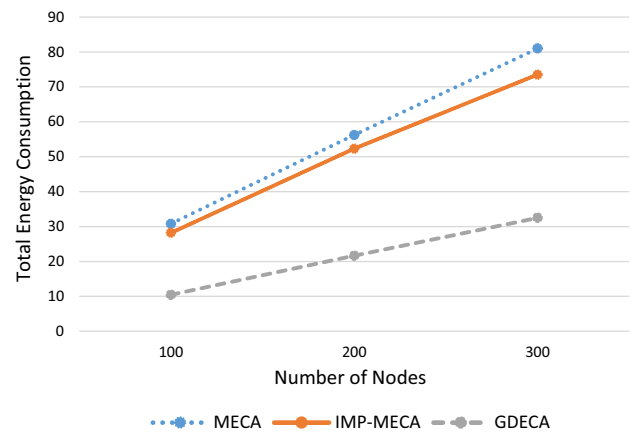
### 5.2.1 Number of mixture models impression on energy consumption

Employing mixture model separates the proposed model from the previous studies. Figure 7 indicates several points in *GDECA*. First, with lower number of GMMs, WSN requires the sinks adjusted to number of GMMs to improve the average distance. For example, with 2 mixtures, employing one sink moving in their intersection reduces average distance. This is also true for 3 GMMs, which requires 3 sinks to cover all intersections. But for 4 GMMs, as number of sinks reduces, average distance decreases, too.

Furthermore, for average residual energy, average residual energy shows that there is little affection on using adjusted number of sinks with same number of GMMs. But generally, with only a single sink to gather data, the residual energy increases. The same analysis is also applied to energy consumption.

Given such discussions, we can conclude that there is trade-off between the number of sinks and energy consummations of nodes to sink. With more sinks to collect data, average distance is reduced, while on the other hand, the number of stoppage stations for all sink is multiplied by the number of sinks. This leads to have an turning point in 3-sinks, where most energy is consumed by nodes.

5.2.2 Number of sinks impression on energy consumption

Number of sinks plays an important role in performance of WSNs and also determining algorithm and policy of a WSN. Figure 8 shows that regardless of WSN size and number of nodes, the performance remains the same. Using 3 mixtures alongside three sinks will obviously reduce the average distance. As we explained in Sect. 5.2.1, using one sink number reduce the total energy consumption and also helps nodes to hold more residual energy. As a result, changing WSN size
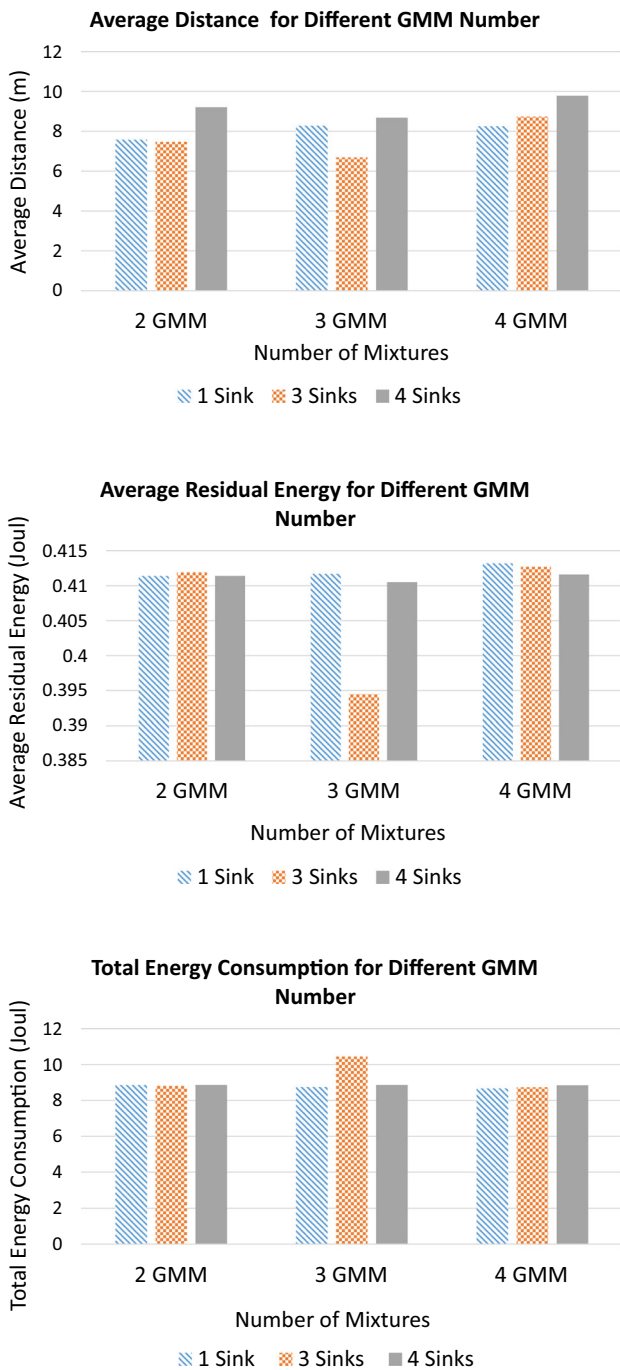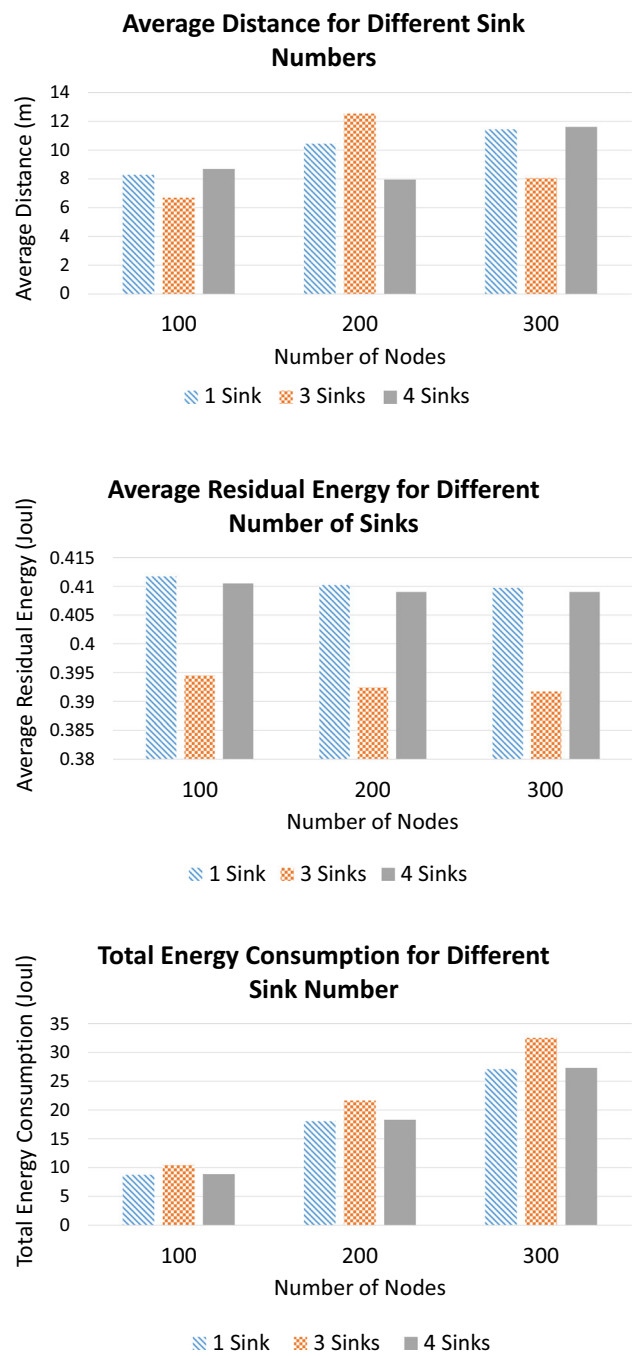
Fig. 7 Performance of *GDECA* regarding to number of mixture models and sink number for 100 nodes

Fig. 8 Performance of *GDECA* regarding to sink number and different number of nodes

will have no effect on WSN performance which helps the proposed approach to be clearly scalable for networks with different number of nodes.

5.2.3 Sink stoppage time impression on performance

In this section, we will examine effects of sink stoppage time on different aspects of the proposed approach. As shown

in Table 2, with increment in stoppage time, average residual energy increases too. This is due to a simple reason: with a fix simulation time, as the stoppage time increases, number of sink stoppage decreases and hence, lower energy will be consumed. On the other hand, total energy consumption decreases. Furthermore, the number of active nodes is reduced as the stoppage time increases. This means that hotspot problem is introduced to WSN when stoppage time

Table 2 Sink stoppage time impression on residual energy, number of active nodes, and total energy consumption on a network with 100 nodes

Stoppage length	Residual energy	Number of active nodes	Total energy consumption
1.2	0.4006	100	9.8446
1.4	0.4061	100	9.2942
1.6	0.4121	99	8.6983

Table 3 Nodes' distribution impression on residual energy, number of active nodes, and total energy consumption on a network with 100 nodes

Nodes' distribution	Residual energy	Number of active nodes	Total energy consumption
Random	0.4006	100	9.8446
Gaussian Distribution	0.3982	99	10.5782
Gaussian Mixtures (2 Mixtures)	0.4104	99	9.3704
Gaussian Mixtures (3 Mixtures)	0.4234	100	7.66

Table 4 Sinks' velocity impression on residual energy, number of active nodes, and total energy consumption on a network with 100 nodes

Nodes' distribution	Residual energy	Number of active nodes	Total energy consumption
3 m/s	0.4391	100	6.09
5 m/s	0.4006	100	9.8446
7 m/s	0.3638	98	14.3476

increases, with no change in sink velocity or number of stations.

5.3 Soft assumption analysis

As mentioned in Sect. 2 we require to design scenarios to evaluate the impact of different soft assumptions. In following, three different scenarios are introduced to evaluate *GDECA*.

5.3.1 Impression of nodes' distribution

In this section, we devised a scenario to analyze the impression of node's distribution on *GDECA* performance. Table 3 represents the performance of proposed model with different distributions on nodes.

The performance is suddenly dropped by using Gaussian model as nodes' distribution, since the proposed model basically fits three mixture models on nodes. Given two and three mixtures of nodes' distributions, the model can fit the desired number of mixtures on nodes and hence the performance of the system is improved, significantly.

5.3.2 Impression of sinks' velocity on GDECA's performance

Previous studies overlook the impression of change in sinks' velocity on the performance of proposed mode. Here we devised a scenario to measure GDECA's performance. As Table 4 shows, the velocity of sinks is strongly contributed to the performance of the model. With velocity increment,

the number of sink stoppage increases over a constant time period. As a result, the more stoppage occurs in sink routing and then more energy is consumed by the nodes and CH.

5.3.3 Impression of number of stations on GDECA's performance

Number of stations plays an important role in overall performance of routing protocols. We examine the impression of number of stations on performance in this section. Table 5 displays the performance of proposed model with different number of stations. As mentioned in Sect. 5.3.2, with more number of stations given for sink, the energy consumption of nodes increases. As a result the average residual energy is decreased.

6 Conclusion

This study introduces a new algorithm; *GDECA*, using Gaussian Mixture Model (GMM), to fit the distribution to nodes, assuming that sensors in real word have mixtures of Gaussian distribution. The proposed approach, helps us to determine mean and variance for each distribution. Using these estimations, we are able to determine *CH* based on distribution means and then we use distributions' intersection to determine sinks' paths. We evaluated proposed approach regarding different metrics such as energy consumption, number of active nodes, number of mixture models, and soft

Table 5 Stations' number impression on residual energy, number of active nodes, and total energy consumption on a network with 100 nodes

Nodes' distribution	Residual energy	Number of active nodes	Total energy consumption
3	0.4219	100	7.81
5	0.4006	100	9.8446
7	0.3912	98	11.6624

assumptions (velocity, nodes' distribution, number of stoppage stations).

Our observations showed that considering nodes distribution is significant in energy saving and also increasing network lifetime. Results showed that average residual energy in nodes are saved by 20–40% and nodes remain active by the end of simulation. In addition, this approach propose a novel idea, which helps nodes to adapt sinks' path based on their distribution and helps the sinks to be adoptable.

Our analysis, also demonstrated that changing sink stoppage time, has no effects on the energy consumption, and conserve the energy with different stoppage station time. Furthermore, different number of nodes makes no change on the performance. This helps the network to be scalable, and keeps the same performance, no matter of its size and number of nodes. Finally, we concluded that different number of mixture models is required to be adjusted with same number of sinks, since sinks' path is determined by clusters' intersections. The best results is obtained when sinks' path is determined by the the number of mixture models, and not randomly.

References

1. Islam, A.K.M.M., Zeb, A., & Wada, K. (2013). Communication protocols on dynamic cluster-based wireless sensor network. In *Informatics, Electronics and Vision (ICIEV), 2013 International Conference on*, pp. 1–6. IEEE.
2. Akyildiz, Ian F., Weilian, Su, Sankarasubramaniam, Yogesh, & Cayirci, Erdal. (2002). Wireless sensor networks: a survey. *Computer Networks*, *38*(4), 393–422.
3. Halder, Subir, & Ghosal, Amrita. (2014). Is sensor deployment using Gaussian distribution energy balanced?. In *2014 IEEE 11th Consumer Communications and Networking Conference (CCNC)*, pp. 721–728. IEEE.
4. Wang, Demin, Xie, Bin, & Agrawal, Dharma P. (2008). Coverage and lifetime optimization of wireless sensor networks with gaussian distribution. *IEEE Transactions on Mobile Computing*, *7*(12), 1444–1458.
5. Yetgin, Halil, Cheung, Kent Tsz Kan, El-Hajjar, Mohammed, & Hanzo, Lajos Hanzo. (2017). A survey of network lifetime maximization techniques in wireless sensor networks. *IEEE Communications Surveys & Tutorials*, *19*(2), 828–854.
6. Faheem, Y., Boudjit, S., & Chen, K. (2011). Dynamic sink location update scope control mechanism for mobile sink wireless sensor networks. In *Wireless On-Demand Network Systems and Services (WONS), 2011 Eighth International Conference on*, pp. 171–178. IEEE.
7. Lin, C.-J., Chou, P.-L., & Chou, C.-F. (2006). HCDD: hierarchical cluster-based data dissemination in wireless sensor networks with mobile sink. In: *Proceedings of the 2006 international conference on Wireless communications and mobile computing*, pp. 1189–1194. ACM.
8. Wang, Jin, Yin, Yue, Zhang, Jianwei, Lee, Sungyoung, & Sherratt, R. Simon. (2013). Mobility based energy efficient and multi-sink algorithms for consumer home networks. *IEEE Transactions on Consumer Electronics*, *59*(1), 77–84.
9. Bishop, Chris. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin: Springer.
10. Boukerche, A., Efstathiou, D., Nikoletseas, S. E., & Raptopoulos, C. (2012). Exploiting limited density information towards near-optimal energy balanced data propagation. *Computer Communications*, *35*(18), 2187–2200.
11. Ebrahimi, Dariush, & Assi, Chadi. (2014). Compressive data gathering using random projection for energy efficient wireless sensor networks. *Ad Hoc Networks*, *16*, 105–119.
12. Narendra, K., & Varun, V. (2014). A comparative analysis of energy-efficient routing protocols in wireless sensor networks. Computer Science and Technology. In *Emerging Research in Electronics* (pp. 399–405). New Delhi: Springer.
13. Chen, S., Coolbeth, M., Dinh, H., Kim, Y.-A. & Wang, B. (2009). Data collection with multiple sinks in wireless sensor networks. Systems, and Applications. In *International Conference on Wireless Algorithms* (pp. 284–294). Berlin, Heidelberg: Springer.
14. Ammari, Habib M. (2013). Joint k-coverage and data gathering in sparsely deployed sensor networks-Impact of purposeful mobility and heterogeneity. *ACM Transactions on Sensor Networks (TOSN)*, *10*(1), 8.
15. Sara, Getsy S., & Sridharan, D. (2014). Routing in mobile wireless sensor network: A survey. *Telecommunication Systems*, *57*(1), 51–79.
16. Ammari, Habib M. (2012). On the problem of k-coverage in mission-oriented mobile wireless sensor networks. *Computer Networks*, *56*(7), 1935–1950.
17. Mario, Di Francesco, Das, Sajal K., & Anastasi, Giuseppe. (2011). Data collection in wireless sensor networks with mobile elements: A survey. *ACM Transactions on Sensor Networks (TOSN)*, *8*(1), 7.
18. Anastasi, G., Borgia, E., Conti, M., & Di Francesco, M. (2011). Reliable data delivery in sparse wsns with multiple mobile sinks: An experimental analysis. In: *Computers and Communications (ISCC), 2011 IEEE Symposium on*, pp. 698–705. IEEE.
19. Zhao, Miao, Ma, Ming, & Yang, Yuanyuan. (2011). Efficient data gathering with mobile collectors and space-division multiple access technique in wireless sensor networks. *IEEE Transactions on Computers*, *60*(3), 400–417.
20. Shi, Lei, Zhang, Baoxian, Mouftah, Hussein T., & Ma, Jian. (2013). DDRP: An efficient data-driven routing protocol for wireless sensor networks with mobile sinks. *International Journal of Communication Systems*, *26*(10), 1341–1355.
21. Lu, Kun-Hsien, Hwang, Shiow-Fen, Yi-Yu, Su, Chang, Hsiao-Nung, & Dow, Chyi-Ren. (2012). Hierarchical ring-based data gathering for dense wireless sensor networks. *Wireless Personal Communications*, *64*(2), 347–367.
22. Van Duc, Le, Oh, Hoon, & Yoon, Seokhoon. (2014). Hicodg: A hierarchical data-gathering scheme using cooperative multiple mobile elements. *Sensors*, *14*(12), 24278–24304.

23. Hu, Yifan, Ding, Yongsheng, Hao, Kuangrong, Ren, Lihong, & Han, Hua. (2014). An immune orthogonal learning particle swarm optimisation algorithm for routing recovery of wireless sensor networks with mobile sink. *International Journal of Systems Science*, *45*(3), 337–350.

24. Liu, Wang, Lu, Kejie, Wang, Jianping, Huang, Liusheng, & Wu, Dapeng Oliver. (2012). On the throughput capacity of wireless sensor networks with mobile relays. *IEEE Transactions on Vehicular Technology*, *61*(4), 1801–1809.

25. Wichmann, A., Chester, J., & Korkmaz, T. (2012). Smooth path construction for data mule tours in wireless sensor networks. In *Global Communications Conference (GLOBECOM)*, 2012 IEEE, pp. 86–92. IEEE.

26. Ha, Ilkyu, Djuraev, Mamurjon, & Ahn, Byoungchul. (2014). An energy-efficient data collection method for wireless multimedia sensor networks. *International Journal of Distributed Sensor Networks*, *10*(9), 698452.

27. Liu, Danpu, Zhang, Kailin, & Ding, Jie. (2013). Energy-efficient transmission scheme for mobile data gathering in wireless sensor networks. *China Communications*, *10*(3), 114–123.

28. Tang, Q., Sun, C., Wen, H., & Liang, Y. (2010). Cross-layer energy efficiency analysis and optimization in WSN. In *Networking, Sensing and Control (ICNSC), 2010 International Conference on*, pp. 138–142. IEEE.

29. Konstantopoulos, Charalampos, Pantziou, Grammati, Gavalas, Damianos, Mpitziopoulos, Aristides, & Mamalis, Basilis. (2012). A rendezvous-based approach enabling energy-efficient sensory data collection with mobile sinks. *IEEE Transactions on Parallel and Distributed Systems*, *23*(5), 809–817.

30. Amini, S. Moloud, Karimi, A., & Shehnepoor, S. R. (2019). Improving lifetime of wireless sensor network based on sinks mobility and clustering routing. *Wireless Personal Communications*, *109*(3), 2011–2024.

31. Pottie, Gregory J., & Kaiser, William J. (2000). Wireless integrated network sensors. *Communications of the ACM*, *43*(5), 51–58.

32. Wahyuni, E. D., & Djunaidy, A. (2016). Fake review detection from a product review using modified method of iterative computation framework. In *Proceeding MATEC Web of Conferences*.

33. Perkins, C. E. (2001). *Ad hoc networking* (Vol. 1). Reading: Addison-wesley.

34. Abdul-Salaam, Gaddafi, Abdullah, Abdul Hanan, Anisi, Mohammad Hossein, Gani, Abdullah, & Alelaiwi, Abdulhameed. (2016). A comparative analysis of energy conservation approaches in hybrid wireless sensor networks data collection protocols. *Telecommunication Systems*, *61*(1), 159–179.

35. Anisi, Mohammad Hossein, Abdullah, Abdul Hanan, Coulibaly, Yahaya, & Razak, Shukor Abd. (2013). EDR: efficient data routing in wireless sensor networks. *International Journal of Ad Hoc and Ubiquitous Computing*, *12*(1), 46–55.

36. Khan, Abdul Waheed, Abdullah, Abdul Hanan, Razzaque, Mohammad Abdur, & Bangash, Javed Iqbal. (2015). VGDRA: a virtual grid-based dynamic routes adjustment scheme for mobile sink-based wireless sensor networks. *IEEE Sensors Journal*, *15*(1), 526–534.

37. Wang, Jin, Zuo, Liwu, Shen, Jian, Li, Bin, & Lee, Sungyoung. (2015). Multiple mobile sink-based routing algorithm for data dissemination in wireless sensor networks. *Concurrency and Computation: Practice and Experience*, *27*(10), 2656–2667.

38. Wang, Jin, Li, Bin, Xia, Feng, Kim, Chang-Seob, & Kim, Jeong-Uk. (2014). An energy efficient distance-aware routing algorithm with multiple mobile sinks for wireless sensor networks. *Sensors*, *14*(8), 15163–15181.

39. Ma, Ming, Yang, Yuanyuan, & Zhao, Miao. (2013). Tour planning for mobile data-gathering mechanisms in wireless sensor networks. *IEEE Transactions on Vehicular Technology*, *62*(4), 1472–1483.

40. Incel, Ozlem Durmaz, Ghosh, Amitabha, Krishnamachari, Bhaskar, & Chintalapudi, Krishna. (2012). Fast data collection in tree-based wireless sensor networks. *IEEE Transactions on Mobile computing*, *11*(1), 86–99.

41. Chen, Young-Long, Wang, Neng-Chung, Shih, Yi-Nung, & Lin, Jia-Sheng. (2014). Improving low-energy adaptive clustering hierarchy architectures with sleep mode for wireless sensor networks. *Wireless Personal Communications*, *75*(1), 349–368.

42. Madani, Sajjad A., Hayat, Khizar, & Khan, Samee Ullah. (2012). Clustering-based power-controlled routing for mobile wireless sensor networks. *International Journal of Communication Systems*, *25*(4), 529–542.

43. Shah, Rahul C., Roy, Sumit, Jain, Sushant, & Brunette, Waylon. (2003). Data mules: Modeling and analysis of a three-tier architecture for sparse sensor networks. *Ad Hoc Networks*, *1*(2–3), 215–233.

44. Medhi, Nabajyoti, & Sarma, Nityananda. (2012). Mobility aided cooperative mimo transmission in wireless sensor networks. *Procedia Technology*, *6*, 362–370.

Houriya Hojjatinia received the M.S. degree in network engineering from Science and Research Branch, Islamic Azad University, Tehran, Iran in 2021. And the B.S. degree in IT engineering from Payame Noor University of Kerman, Iran, in 2012. Her research interests are Wireless Networks, Communication systems And robust routing algorithms.
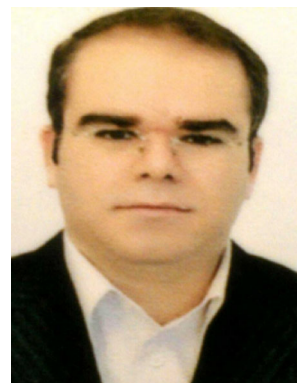
Mohsen Jahanshahi is currently dean of technical faculty in Islamic Azad University, central Tehran branch. His main research Interests are Wireless Networks, Mathematical Optimization, and Soft Computing.

Saeedreza Shehnepoor received his BS from Sharif University of Technology, at 2014 in Information Technology (IT). He finished his master in Artificial Intelligence and received MS degree from University of Tehran (UT) at 2016. He is currently a second year PhD student in University of Western Australia (UWA). His research field is concentrated on Machine Learning area and specifically issues related to the fraud contents in social media.