Human-centric Computing
and Information Sciences

Open Access

CrossMark

An energy-efficient sleep scheduling mechanism with similarity measure for wireless sensor networks

Runze Wan[1], Naixue Xiong[2*] and Nguyen The Loc[3]

*Correspondence:
xiongnaixue@gmail.com
[2] Department
of Mathematics
and Computer Science,
Northeastern State University,
Tahlequah, OK, USA
Full list of author information
is available at the end of the
article

Abstract

In wireless sensor networks, the high density of node's distribution will result in transmission collision and energy dissipation of redundant data. To resolve the above problems, an energy-efficient sleep scheduling mechanism with similarity measure for wireless sensor networks (ESSM) is proposed, which will schedule the sensors into the active or sleep mode to reduce energy consumption effectively. Firstly, the optimal competition radius is estimated to organize the all sensor nodes into several clusters to balance energy consumption. Secondly, according to the data collected by member nodes, a fuzzy matrix can be obtained to measure the similarity degree, and the correlation function based on fuzzy theory can be defined to divide the sensor nodes into different categories. Next, the redundant nodes will be selected to put into sleep state in the next round under the premise of ensuring the data integrity of the whole network. Simulations and results show that our method can achieve better performances both in proper distribution of clusters and improving the energy efficiency of the networks with prerequisite of guaranteeing the data accuracy.

Keywords: Wireless sensor networks, Load balance, Sleep scheduling, Energy-efficient

Introduction

In recent years, there has been a great deal of interest researches in wireless sensor networks (WSNs), which involves the maturing techniques of integrated circuitry, micro electro mechanical systems (MEMS) and digital signal processing [1]. WSNs usually can be composed of hundreds or thousands of sensor nodes, each one of which is capable of sensing its environment, performing simple computations, and communicating to its neighbors [2]. Due to the sensing behavior on a larger geographical region and sending their readings (raw data) to the sink, the sensor nodes with limited energy being supplied from built-in battery will die gradually. Apparently, the power saving of sensor nodes is vital to prolong the lifetime of the entire network.

In order to gather information more efficiently, hierarchical cluster-based structure is introduced into the applications of WSNs [3]. In such scenario, the sensor nodes can be partitioned into a number of small groups called clusters. Each cluster is composed of a coordinator as cluster head (CH), which is responsible for managing the entire cluster

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 2 of 22

and forwarding the data to the base station (BS). By rotating cluster-heads selection periodically, the node's energy consumption over the network can be balanced. In case of dense-deployment, the readings being collected by sensor nodes in the adjacent regions may demonstrate the features with spatial and temporal correlations. Despite of providing a fault-tolerant mechanism for data aggregation, the redundant data will result in superfluous data transmission, and it will lead to collisions and undesired energy depletion to affect the network lifetime [4]. In view of the recognizable targets, the evolutionary algorithms are proposed to arrange the nodes with similar monitoring results into the same cluster as far as possible. The amount of data that the CHs communicate with the BS is maximally compressed by the fusion of similar information. Finally, the sensor nodes with spatial-correlation can be organized as much as possible in a cluster. Therefore, it can not only improve the accuracy of the data in the monitoring area, but also reduce the transmission cost of the CH. However, this kind of methods has high complexity and long time consuming [5, 6].

In this paper, we propose an energy-efficient sleep scheduling mechanism with similarity measure for WSNs (ESSM), which will schedule the sensors into active or sleep mode to reduce energy consumption effectively. Firstly, the optimal competition radius is estimated to organize the all sensor nodes into several clusters to balance energy consumption. Secondly, according to the data collected by member nodes, a fuzzy matrix can be obtained to measure the similarity degree, and the correlation function based on fuzzy theory can be defined to divide the sensor nodes into different categories. Next, the redundant nodes will be selected to put into sleep state in the next round under the premise of ensuring the data integrity of the whole network.

The rest of this paper is organized as follows: In "Related work" section, we briefly introduce related work. We describe the assumptions and explain the details of our method in "System model and analysis" section. In "Sleep scheduling algorithm" section, the detailed of sleep scheduling algorithm is described. At next Section, the experiments method is shown and the result is discussed regarding the performance evaluation of our method. Finally, we conclude this paper and discuss the future work in "Conclusions and future work" section.

Related work

At present, the cluster organization is widely used in WSNs, and it has been proved to be a beneficial way to improve energy efficiency and extend the network lifespan. Heinzelman et al. [7] proposed low-energy adaptive cluster hierarchy (LEACH) protocol for wireless micro-sensor networks. In that manner, the sensor nodes being situated in different regions are organized into several clusters, and the CHs should be selected to aggregate data from their respective member nodes and forward to the BS.

Subsequently, there have been several studies for reducing data traffic in cluster-based WSNs by exploiting the correlations among readings of sensor nodes. Kumar et al. [8] proposed an energy-efficient heterogeneous clustered scheme (EEHC) for WSNs, in which a percentage of nodes are equipped with more energy than others and the nodes play the role of a CH based on the weighted election probabilities according to the residual energy. Karaca et al. [9] proposed analytic hierarchy process (AHP), which is applied to centralized CH selection. They discussed the factors that

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 3 of 22

contribute to the partition of clusters, including residual energy and the distance to the cluster's centroid. Hou et al. [10] proposed an improved LEACH Protocol in the aspect of selection of the CHs and energy consumption balance. Wu et al. [11] introduced a dynamic sleep scheduling mode rather than the fixed sleep/wake-up mode in LEACH, which can improve the network lifetime remarkably. In [12], we investigated the above problems and proposed an unequal clustering mechanism for inter-cluster multi-hop routing, which divides all nodes into clusters with unequal size. The clusters closer to the BS will be designated as smaller size. Therefore, the CHs of those clusters can preserve some more energy for the inter-cluster relay traffic, and the "hot-spots" problem can be alleviated effectively. Sajjanhar et al. [13] proposed a distributive energy efficient adaptive clustering protocol (DEEAC), which owns spatio-temporal variations in data reporting rates across different regions and selects the sensor node to be a CH depending upon its hotness value and residual energy. Ali et al. [14] proposed advanced LEACH routing protocol for WSNs, which combines the current state probability and general probability to CH's selection in each round.

For clustered WSNs, itinerary-based R-tree (IR-tree) have been presented to solve the existing problems of spatial query processing [15], which make efficient processing and expansion of the query due to the spatio-temporal characteristic of data samples in each cluster. By extracting the principal components of spatially correlated sensor data collected from member nodes, a cluster-based data analysis framework is proposed to aggregate the redundant data as well as detect the outliers in the meantime [16]. In [17], an energy efficient hybrid node scheduling scheme (EEHS) in cluster-based WSNs is proposed to improve the overall efficiency and network lifetime, which can identify the nodes with redundant coverage in each round. In [18], a sleep scheduled and tree-based clustering routing protocol for energy-efficient WSNs (SSTBC) is presented to preserve the energy of entire network by turning off radio of either impossible or unnecessary nodes. In addition, it builds minimum spanning tree with the root as the CH for data forwarding, which can reduce energy dissipation for long distance transmission.

Spatial correlation can be expressed as data similarity between adjacent nodes in a region. The nodes with adjacent spatial correlation often lead to massive redundancy in data transmission. Accordingly, the common technique is to put some sensors in the sleep or active mode dynamically for sensing and communication. It can save energy by scheduling the sensors into appropriate state, and also reduce communication range and control messages. In [19], a sleeping scheduling scheme based on AHP is proposed. Some factors, such as the distance to CH, the residual energy, and sensing coverage ratio are taken into account to achieve the optimal nodes scheduling decision. In [20], a cross-layer organizational approach based on sleep scheduling is proposed, which can increase the monitoring coverage as well as the operational lifetime in wide-area surveillance applications. By defining the residual energy level of the nodes, More et al. [21] proposed a random backoff sleep protocol (RBSP) to increase the network lifetime and balance the energy consumption among the nodes. RBSP can ensure that the probability of neighbor nodes becoming active is inversely related to the residual energy level of the current active node. They also proposed Optimized Discharge-Curve-based Coverage Protocol (ODCP) to resolve the problem of coverage gaps within the sensing area as the sudden failure of nodes happens [22]. Their scheme can determine optimal sleep schedules for

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 4 of 22

redundant nodes using their neighboring active nodes' battery discharge rate, failure probability, and coverage overlap information.

Some methods [23–25] make use of statistical models to estimate the readings of nodes. Owing to require few readings to respond of queries, statistical models can drastically reduce the amount of data sent by nodes [26]. However, the inadequacies lie in time cost and energy consumption being required for constructing those models in terms of massive data. In addition, they are unable to decrease the conflict and interference originated from the redundant nodes.

System model and analysis

Network model

There are N sensor nodes being deployed randomly with a square field and the BS is located far away from square sensing field. The nodes observe the area continuously and transmit the observation results to the BS periodically. s_i represents the i-th sensor node, and the corresponding sensor nodes set is $S = \{s_1, s_2,..., s_N\}, |S| = N$. To determine the optimal parameters for our model, we make the following assumptions:

1. The BS is located outside the square area and far away from the observation field. Once being deployed, the BS and sensor nodes will keep stationary.
2. The location-unaware sensor nodes are uniformly distributed in the observation area, and each of them will be allocated with a unique ID.
3. After detecting the strength of the signal, the sensor node can estimate the approximate distance from the sender, and adjust the transmission power adaptively to save energy according to the distance.
4. All sensor nodes are capable of data fusion. Assuming data being perfect relevant, they can be fused into a plurality of packets with equal size.

In this paper, the energy consumed during per round can be estimated based on the energy consumption of the nodes for reception or transmission in each round. To measure the energy consumption, the First Order Radio model is exploited [7]. The energy spent for transmission of l-bit packet from the transmitter to the receiver at a distance (d) can be defined as:

$$E_{Tx} = \begin{cases} lE_{elec} + l\varepsilon_{fs}d^2, & d < d_0 \\ lE_{elec} + l\varepsilon_{mp}d^4, & d \geq d_0 \end{cases} \tag{1}$$

where E_{elec} is the dissipated energy to operate the transmitter or receiver circuitry per bit, d is the transmission distance. ε_{fs} and ε_{mp} are the amplifier energy factors for free space and multi-path fading channel models, respectively. The cross over distance d_0 is the threshold that depends on the specific scene and the amplifier energy factors, which can be given as $d_0 = \sqrt{\varepsilon_{mp}/\varepsilon_{fs}}$.

The energy being spent for receiving l-bit data can be given as:

$$E_{Rx}(l, d) = lE_{elec} \tag{2}$$

and the energy consumption for data aggregate by CH is:

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 5 of 22

$$E_{Aggr}(l, d) = lE_{DA} \tag{3}$$

where E_{DA} is the energy consumption of data fusion per unit.

Optimal competition radius

In order to prolong the lifetime of the network, it is necessary to balance the energy consumption among sensor nodes [27]. For reducing the energy consumption in the cluster, the inter-cluster communication distance should be restricted within the threshold d_0, and it will ensure to keep in touch of the sensor's energy loss by free space model. Under the condition of the single hop mode, the CH can send data to the BS directly. If the BS is far away from the monitoring area, the CH needs to employ multi-path attenuation model to deal with the power amplification loss. It will increase the energy consumption of CH greatly. Therefore, CHs are more likely to consume energy and turn to die earlier than its member nodes, which will shorten the lifetime of the whole network badly. Therefore, it is crucial to set optimal competition radius of CHs and form distributed and uniform clusters to balance the energy consumption.

When a node broadcasts a candidate-CH message, the range is termed as the competition radius of candidate-CH. Only the nodes within the radius of the competition can receive the message from the candidate CH. During the stage of the CH selection, the spatial distribution of CHs can be restricted by setting the competition radius. In a single hop mode, the energy of the sensor node may be primarily used for sending data to the BS. Therefore, the radius of competition can be regarded as the key parameter for affecting the lifetime of the network. If the radius value is larger, it will result in a lower number of clusters, and more energy consumption due to the higher signal power with respect to transmit across a larger distance. In brief, the appropriate competitive radius will be conducive to balance the energy consumption of CHs and the overhead of intra-cluster communications.

Suppose that the monitoring area is a square with the length M and N sensor nodes are deployed, then the node's density is $\rho = N/M^2$. Let d_{CH} denote the competition radius of CH, d_{toBS} is the transmission distance from the CH to BS. Some redundant nodes will be scheduled into sleep state, and the detailed mechanism will be discussed in "Determination of redundant nodes" section. Suppose that the number of redundant nodes in the cluster is V, the percentage of active member nodes in the cluster member node will be $\sigma = 1 - V/(n-1)$. For simplicity, the length of the packet being delivered to CH from the active member node is set to l-bits in each round. E_{toCH} indicates the energy consumption for sending l-bits packet from the member node to its CH, E_{re} is the energy consumed by CH for receiving such packet, and E_{toBS} denotes the energy consumption for sending l-bit packet from the CH to BS. In addition, the energy consumption of all active nodes with range of CH's competition radius to transmit their collected data to the CH can be obtained:

$$
\begin{aligned}
E_{toCH} &= \sigma \int_{d_{CH}}^{0} 2\pi x \times \rho \times \left(lE_{elec} + l\varepsilon_{fs}x^2 \right) dx \\
&= \sigma l\pi\rho \times \left(E_{elec}d_{CH}^2 + \frac{1}{2}\varepsilon_{fs}d_{toBS}^4 \right)
\end{aligned} \tag{4}
$$

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 6 of 22

where $2\pi x \times \rho \times dx$ is the number of nodes being located in the ring annular area with the length $x(0 \leq x \leq r_i)$.

For each CH, the energy consumed for receiving monitored data from its member nodes can be estimated as:

$$E_{re} = l \times \left(\pi r_i^2 \rho - 1 \right) \times E_{elec} \tag{5}$$

The energy consumed of CH for data aggregation is given as:

$$E_{ag} = l \times \pi r_i^2 \rho \times E_{DA} \tag{6}$$

Besides, the energy consumed by CH for data forwarding to the BS can be given as:

$$E_{toBS} = lE_{elec} + l\varepsilon_{mp}R_i^4 \tag{7}$$

where R_i is the distance from the CH to the BS.

Therefore, the total energy dissipation in a cluster can be calculated as:

$$E_{cluster} = E_{toCH} + E_{toBS} + E_{re} + E_{ag} = l\pi\rho r_i^2 \left(\sigma + \frac{1}{2}\sigma\varepsilon_{fs}r_i^2 + E_{elec} + E_{DA} \right) + l\varepsilon_{mp}R_i^4 \tag{8}$$

Then, the average energy consumed by a node in a cluster is:

$$E_{avg} = \frac{E_{cluster}}{\pi r_i^2 \rho} \tag{9}$$

By taking the derivative of r_i in formula (9), we can derive the optimal competition radius d_{CH} as follows:

$$d_{CH} = \sqrt[4]{\frac{2\varepsilon_{mp}}{\sigma\pi\rho\varepsilon_{fs}}} d_{toBS} \tag{10}$$

From the formula (4) and (5), it can be observed that the optimal competitive radius of each node will increases as the distance between the node and the BS. According to the optimal competition radius of each node, it can minimize local energy consumption, and come into being an uneven hierarchical structure of clustered WSNs. In the region close to the BS, the distribution density of the clusters will have relatively small size.

Selection of CHs

During the phase of CH-determination, CHs will be determined based on a linear combination of probability selection and local competition. Once the operation is complete, all nodes may be put into the status of member, candidate-CH or CH. Initially, each node is treated as common node, and will generate a certain probability for being candidate-CH in view of the distance from the BS and the residual energy. Comparing with the nodes far away from the BS, the sensors that are close to the BS will have higher probability of being candidate-CHs. Therefore, the probability function can be defined as follows:

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 7 of 22

$$CHS(i) = \alpha \times \frac{d_{\max} - d_{toBS}(i)}{d_{\max} - d_{\min}} + (1 - \alpha) \times \frac{E_{res}(i)}{E_{init}} \qquad (11)$$

where $d_{\max}$ denotes the is the largest distance to the BS, and $d_{\min}$ denotes the nearest distance. α is the constant parameter.

Each node will obtain the probability of being candidate-CHs to ascertain whether it has competencies required to be a CH in current round. Moreover, the nodes being candidate-CHs will modify their own state, and calculate the optimal competitive radius according to their distance to the BS. To save energy, other nodes that fail to become candidate-CHs can turn off the module of wireless communication during the phase of CH selection.

Besides, the candidate-CHs should acquire the location information and adjacent competitors within its communication range. Each candidate-CH maintains a neighbor candidate-CH table, which includes the node's ID, the remaining energy and status flag of adjacent CHs. After CH's selection, CHs will broadcast the message, which includes their identity and the member nodes list, and wait for adjacent member nodes to join in. According to the received signal strength, non-CH node estimates the distance of its neighboring CHs, and choose to join in the nearest one. It can reduce the energy consumption of the member nodes for data delivery, and also make the nodes near the BS undertake more burden of data forwarding or aggregation to achieve balanced energy consumption of the entire network. Next, the member nodes will send to the CH with the join_MSG, which contains the node's ID and distance between the CHs. By receiving the join_MSG from the non-CHs, the CH will send ACK message and update the cluster membership list simultaneously.

Determination of redundant nodes

Since the member nodes are often distributed in the adjacent region, the data collected by those sensors tend to demonstrate spatial and temporal correlations [27]. In order to reduce unnecessary energy consumption, some sensor nodes maybe turn into the sleep state. Our work mainly focuses on the node's sleep schedule, which is generally more suitable to solve the problems of data redundancy and transmission conflict. Moreover, by making some nodes enter the dormant state, the sleep scheduling strategy can save the energy consumption derived from node's active state. In this section, we will discuss how to extend the WSN's lifetime by using optimization strategy based on fuzzy clustering theory.

The main idea is based on classification by fuzzy mathematics, which can group the nodes with high data similarity into the same category [28]. According to the scheduling mechanism, a certain mount of sensor nodes will be selected from all categories. More concretely, while the perceptual data received by CH from its member nodes is accumulated to a certain extent, the fuzzy similarity matrix can be constructed to make clustering. Next, in the premise of the data fusion accuracy as high as possible, some nodes can be chosen from all categories as redundant ones. Finally, a specific sleep scheduling mechanism can be applied in those redundant nodes to reduce the communication costs and traffic conflicts.

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 8 of 22

The mutual support degree between node s_i and s_j can be defined as the confidence distance, which is expressed as function *Del(i, j)*. The smaller the value is, the closer the measurement value of the pair of sensor nodes is. Conversely, it demonstrates that the monitored data collected by those nodes differ considerably. Clustering method is based on fuzzy matrix to classify the observed objects. For different confidence levels, different classification results can be obtained, and then to form a dynamic clustering diagram.

Assuming domain $S = \{s_1, s_2, ..., s_n\}$ denotes the set of member nodes of a cluster, and n is the number of nodes in the cluster. The monitoring time can be divided into m intervals, and x_{ij} represents the data collected by the member node s_i at time j. The original data matrix can be given as $X = (x_{ij})_{n \times m}$.

After standardization, the matrix X will be transformed into a fuzzy matrix [29]. Firstly, the shift and standard deviation transformation is implemented and the element in the normalized matrix can be given as

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{z_j}, \quad (i = 1, 2, \ldots, n, \ j = 1, 2, \ldots, m) \tag{12}$$

where $\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}$, $z_j = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}$, $(j = 1, 2, ..., m)$.

For $x'_{ij} \notin [0, 1]$, it is necessary to make further process for the point of view of unified dimension.

$$x''_{ij} = \frac{x'_{ij} - \min_{1 \le i \le n}\{x'_{ij}\}}{\max_{1 \le i \le n}\{x'_{ij}\} - \min_{1 \le i \le n}\{x'_{ij}\}}, \quad (j = 1, 2, \ldots, m) \tag{13}$$

Thus, the fuzzy similar matrix $R = \left(x''_{ij}\right)_{n \times m}$ can be obtained.

Taking into account that temporal correlation of the data collected by those nodes, the correlation coefficient method is employed to form the fuzzy similarity matrix.

$$r_{ij} = \frac{\left|\sum_{k=1}^{m} (x_{ik} - \bar{x}_j)(x_{jk} - \bar{x}_j)\right|}{\sqrt{\sum_{k=1}^{m} (x_{ik} - \bar{x}_i)^2}\sqrt{\sum_{k=1}^{m} (x_{jk} - \bar{x}_j)^2}} \tag{14}$$

Next, the λ-truncation matrix $R_\lambda = (r_{ij}(\lambda))$ can be deduced according to the fuzzy similarity matrix. Since R_λ is Boolean matrix, the classification of the sensor nodes depends on whether the element value of matrix R_λ is equal to 1 or not. The specific principles are as follows: (i) If both R_λ and R are equivalent matrixes, the nodes can be categorized directly; (ii) Otherwise, until R_λ being converted into an equivalent Boolean matrix by certain rules, the classification method will not be implemented.

Furthermore, the fuzzy similarity matrix is used to obtain the clustering graph. By choosing $\lambda_1 = 1$ and generating the equiform class $[x_i]_R = \{x_j | r_{ij} = 1\}$ for each x_i, the node x_j can be attributed into the same class when the condition is satisfied. By taking

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 9 of 22

$\lambda_2(\lambda_2 < \lambda_1)$ as the second maximum value, the elements pair (x_i, x_j) with the similarity degree λ_2 can be found out from the matrix R, i.e. $r_{ij} = \lambda_2$. Similarly, by merging x_i and x_j in the equivalent classification λ_1 into one class, the equivalent classification on the level λ_2 can be obtained. And so forth, let $\lambda_1 > \lambda_2 > \cdots > \lambda_k$ until S is merged into a single class, the number of k categories for clustering can be obtained.

After clustering, the sensor nodes in same cluster can be partitioned into several categories based on the similarity of monitoring data. In each category, some nodes may be selected as redundant nodes and scheduled to sleep state. It will reduce the overall energy consumption of the network significantly. Obviously, if the number of redundant nodes selected is less, the greater the amount of information is retained on the whole, that is, the data collected is more comprehensive.

Let $s_i^{(v)}$ denote the i-th node in category v, and the number of nodes in category v can be expressed as $V = |s^{(v)}|, \sum_{v=1}^{k} |s^{(v)}| = n$. To measure the difference between the data being collect at time m, we have

$$Del(s_i^{(v)}, s_j^{(v)}) = \sqrt{\sum_{u=1}^{m} (x_{iu}^{(v)} - x_{ju}^{(v)})^2}, \quad (i, j = 1, 2, \ldots, V, \; i \neq j) \tag{15}$$

According to the principle of redundant nodes' selection, it should be guaranteed to minimize the amount of information loss. Thus, the objective function for selecting redundant nodes can be given as:

$$s_*^{(v)} = \arg\min \left\{ \sum_{i=1}^{V} Del(s_i^{(v)}, s_j^{(v)}) \right\} \tag{16}$$

Eventually, $s_*^{(v)}$ denotes the redundant nodes selected from the category v.

Sleep scheduling algorithm

In this section, the proposed energy-efficient sleep scheduling mechanism (ESSM) is described in detail. ESSM is a distributed competitive mechanism based on unequal cluster-WSN, and it makes local decisions for determining competition radius and electing cluster-heads. In order to estimate the competition radius for tentative cluster-heads, ESSM employs both residual energy and distance to the BS parameters. Moreover, it takes advantage of fuzzy logic method to acquire optimal competition radius based on a probabilistic model, which is employed for competition between candidate cluster-heads. The specific flow chart of ESSM is shown in Fig. 1.

The candidate-CHs broadcast their participation information to adjacent nodes in the range of competition radius with the corresponding transmission power. The comp_MSG message contains the candidate-CH ID and residual energy. If receiving that message, the other candidate-CHs will record the ID of the candidate-CH to the neighbor candidate-CH table. Due to different size of the competition radius from candidate-CHs,

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 10 of 22

Fig. 1 Flow chart of ESSM
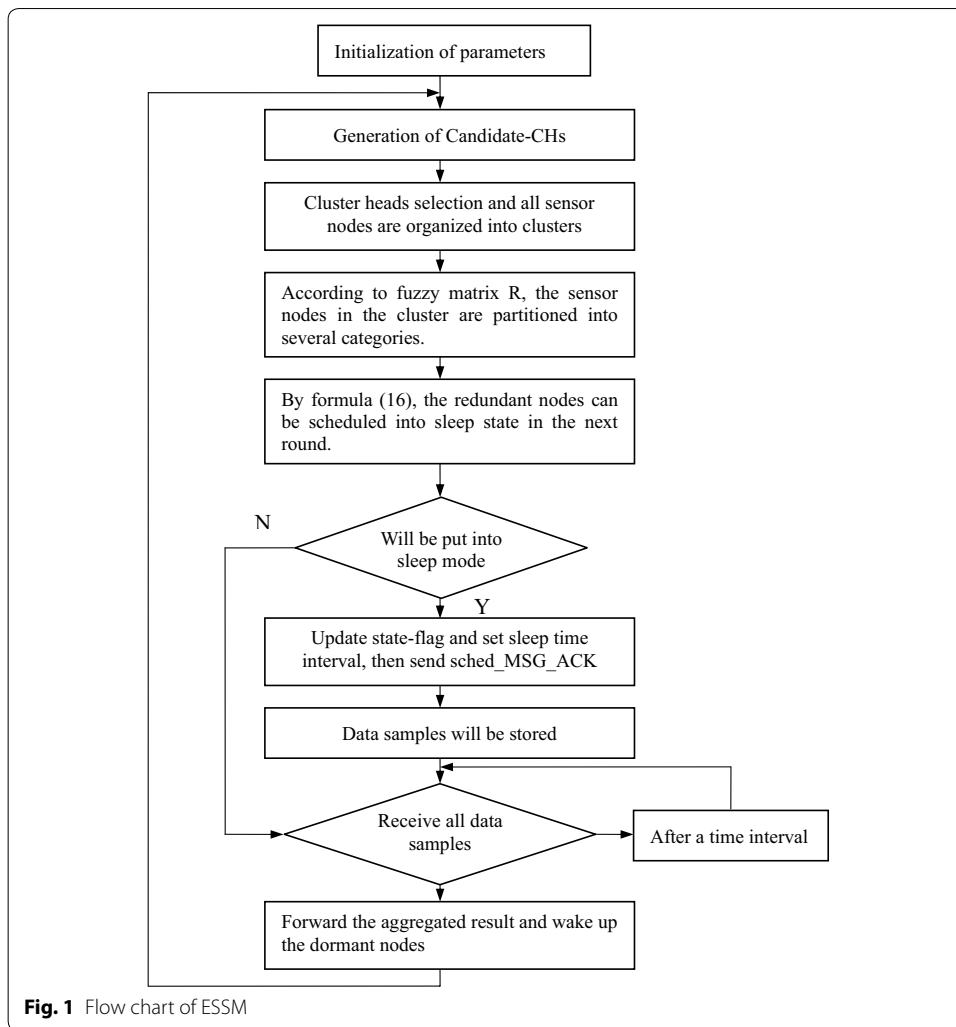
the following case may occur. Assuming that competition radius of candidate-CH s_i is larger than candidate-CH s_j and s_j can receive the comp_MSG message from s_i. However, the comp_MSG message from s_j can not be transferred to s_i due to the limitation of the transmission range of s_j, and it will result in that s_i not be aware of the existence of candidate-CH s_j. In order to acquire complete information about adjacent competitors, any candidate-CH must estimate the distance from the sender after receiving the comp_MSG message. If the distance is greater than the radius of their competition, it is necessary to reissue a comp_MSG message to the sender. In this way, it can assist the senders in acquiring complete neighbor candidate-CH information and updating the neighbor candidate-CH table.

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 11 of 22

After receiving the comp_MSG message, candidate-CH will compare its residual energy with the sender's. If the remaining energy is less than sender's, it will quit the competition, set its status as member node, and broadcast quit_MSG. Otherwise, it will wait for the comp_MSG from other competitors till the end of the CH selection. If still does not withdraw from the competition, it will broadcast sus_MSG message to all nodes in its transmission range to declare being elected to the CH, and modify the state flag. If a candidate-CH receives the quit_MSG, the node will check whether the final state has been determined or not. If already being the CH or member, it will drop this message. Otherwise, once the node is still of the candidate-CH state, it will update the status of the sender to member state in the neighbor candidate-CH table, and continues to wait for the messages from other neighbor candidate-CHs to decide its final state.

In the process of data delivery, the members make use of TDMA mechanism to send data to CH. First, the CH will divide data collection time into several time slots, and attribute each interval to a member node to form a scheduling arrangement for data aggregation. Then, according to the distance of its member nodes, the CH can set its transmission range and send sched_MSG to all member nodes for scheduling. After receiving the sched_MSG, the member nodes record it belonging time interval for data transmission according to the sched_MSG respectively. For saving energy, they can turn off the wireless communication unit during the non-transmission slots.

While the CH receives the data from all the member nodes, it will restore them and further make use of the fuzzy clustering method to select redundant nodes. Next, the nodes in redundant nodes set $\{R_1, R_2,..., R_t\}$ will be notified, and they will modify the state flag and be scheduled to dormant state in the next round. After that, the CH will incorporate data aggregation to reduce the amount of information, and send the aggregation result to the BS in sing-hop manner. Finally, the BS receives all messages sent by the CH, and end the round. By the end of each round, all nodes will update the status as member nodes and empty their neighbor CHs tables. Table 1 illustrates the messages definition, and the main steps are explained in detail.

Table 1 Messages definition

Message	Description
comp_MSG	Contains the candidate CH ID and residual energy
temp_CH_msg	Contains neighbor candidate CH information
sus_MSG	Nodes succeeding to be elected as CH send this message
quit_MSG	Nodes failing to be elected as CH send this message
join_ MSG	Member nodes search for the closest CH to ask for joining
TDMA_msg	CHs allocate time intervals to member nodes and form a scheduling arrangement
mornitor_MSG	Packages include CH identity, sequence and sampled data
mornitor_MSG_ACK	Acknowledgement for *mornitor_MSG*
attention_MSG	Attention message to member nodes
sched_MSG	Notify the nodes being scheduled to dormant state in the next round
sched_MSG_ACK	Acknowledgement for *sched_MSG*

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 12 of 22

Step 1: Generation of CHs.

BS broadcast ADV;

for *each node* s_i

   Count the distance $d_{toBS}(i)$ *based on RSSI;*

   According to $d_{\max}$, $E_{res}(i)$ *and initial energy* E_0, *Count* $CHS(i)$;

   Random$(0,1) \rightarrow t_0$;

   if *(* $t_0 < CHS(i)$ *)*

      Add node s_i *to candidate-set-CH;*

$$d_{CH}(i) = \sqrt[4]{\frac{2\varepsilon_{mp}}{\sigma\pi\rho\varepsilon_{fs}}} d_{toBS}(i) ;$$

      Broadcast a comp_MSG(ID, $E_{current}$, d_{CH}, d_{toBS} *) to other nodes;*

         Set $d_{CH}(i)$ *to the optimal competition radius for node* s_i;

   else *exit;*

   end if

end for

for *each node* $s_j \in neighbor\_SET(s_i)$

   while *receive temp_CH_msg from node* s_i;

      Count d_{ij};

      if $s_j \in$ *candidate-set-CH &&* $d_{ij} \leq \min\{d_{CH}(i), d_{CH}(j)\}$

         Add node s_j *to set_CH;*

      else

         if $E_{res}(j) > E_{res}(i)$

            s_j *broadcast sus_MSG(ID);*

         else

            s_i *broadcast quit_MSG(ID);*

         end if

      end if

   end while

end for

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 13 of 22

Step 2: Determination of member nodes.

for *each node in state of member node*

    if *receive sus_MSG(ID)*

      Send join_ MSG(ID, id_CH) to corresponding CH according to RSSI ;

    else

      Receive TDMA_msg(ID, id_CH, Timeslot);

      Waiting for data transfer;

    end if

end for

Step 3: Intra-cluster data transmission.

for *each CH*

    while *(i<m)*

        for *each Timeslot*

            if *receive mornitor_MSG(id_CH, sequence, data);*

                Send mornitor_MSG_ACK;

            else

                Send attention_MSG(id_CH, id, sequence_num);

            end if

        Interval(i)++;

    end while

end for

for *each non-CH*

    send mornitor_MSG(ID, sequence, data);

    wait;

    receive mornitor_MSG_ACK;

end for

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 14 of 22

Step 4: Determine the node to be scheduled to sleep.

for *each CH*

  Construct Raw data matrix X;

  After the data is standardized, it is transformed into a fuzzy matrix R;

  The member nodes are divided into k categories;

  for *each category v*

    for *each node j, k$\in s^{(v)}$*

      calculate $Del(s_i^{(v)}, s_j^{(v)})$ between the data from node $s_i^{(v)}, s_j^{(v)}$;

$$s_*^{(v)} = \arg\min\{\sum_{i=1}^{V} Del(s_i^{(v)}, s_j^{(v)})\};$$

    end for

  end for

  Obtain redundant nodes set $\{R_1, R_2, \ldots\ldots R_t\}$;

  for *each node$\in \{R_1, R_2, \ldots\ldots R_t\}$*

    send sched_MSG(id_CH, id, state_flag);

    receive sched_MSG_ACK;

  end for

end for

Experimental results and analysis

We implement the sleep scheduling mechanism in a custom WSN simulator build in C++, and conducted several experiments to evaluate its performance. In this study, ESSM was compared with two protocols, EEHS [17] and SSTBC [18]. The detailed simulation parameters are illustrated in Table 2.

First, we examine the effect of parameter α on the balance of network energy consumption. The ratio of node's sleep scheduling can illustrate the energy saving effect, but it can not really reflect the balance of network energy consumption. In general, the metric of mean residual energy and variance of all sensor nodes' at a certain time can be applied to measure the balance of network energy consumption. It can be observed that the greater the mean energy, the lower variance of all sensor nodes and the better the consumption balance of the mechanism can be obtained. The energy mean function can be defined as $M(t) = \sum_{i=1}^{N} E_i(t)/N$. Moreover, The energy variance function of the network can be given as $D(t) = \sum_{i=1}^{N} [E_i(t) - M(t)]/N$. The setting of the probability function takes into account of the residual energy and the distance from nodes to BS.

The experimental results are shown in Figs. 2 and 3. The values of the parameter α vary from 0.2, 0.5 and 0.8. It can be seen that when the value is 0.5, the result of energy mean and variance are better than other cases. That is because the smaller the
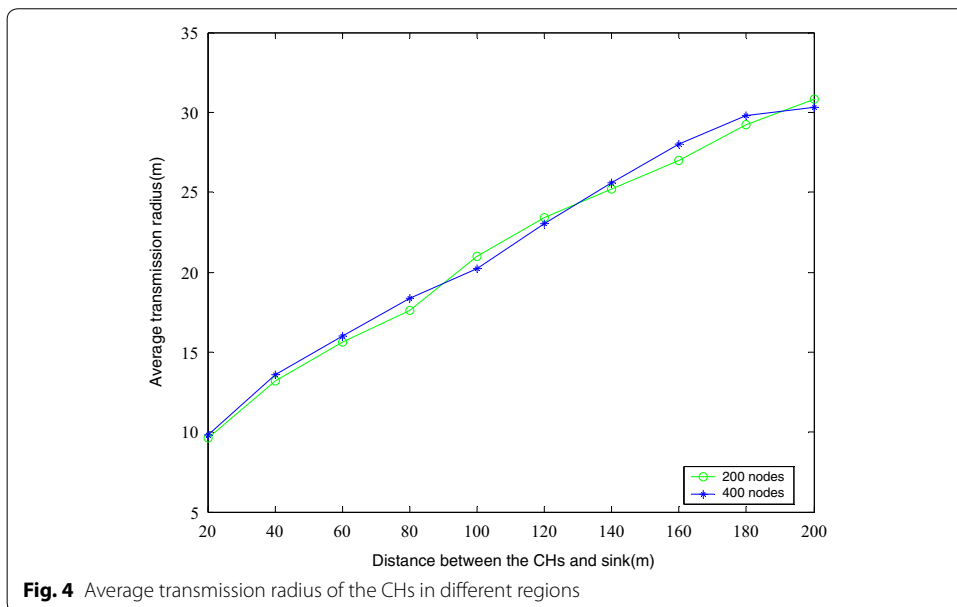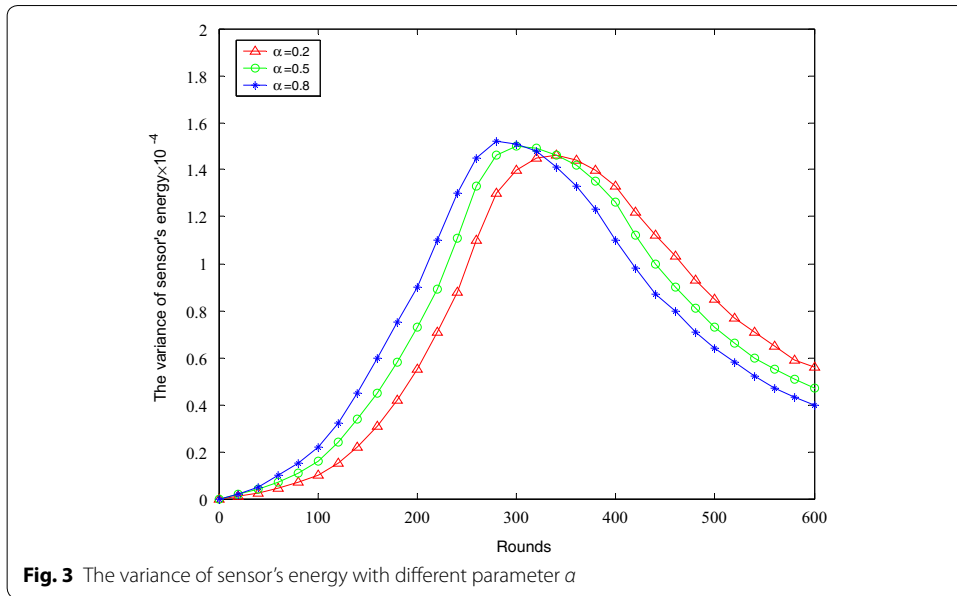
Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 15 of 22

Table 2 Simulation experiment parameters

Parameters	Value
Network interface	Channel/wireless
Packet size (bytes)	512
Network size	200×200 m
BS location	(200, 100) m
Round duration	10 s
Initial energy in each node	0.1 J
Number of nodes	100
% of threshold for residual energy	0.35
Data rate	1 kbps
ε_{mp}	0.0013 pJ/bit/m^4
ε_{fs}	10 pJ/bit/m^2
E_{elec}	50 nJ/bit
E_{DA}	5 nJ/bit/signal
Idle power	13.5 mW
Sleep power	15 lW
Threshold distance (d_0)	75 m
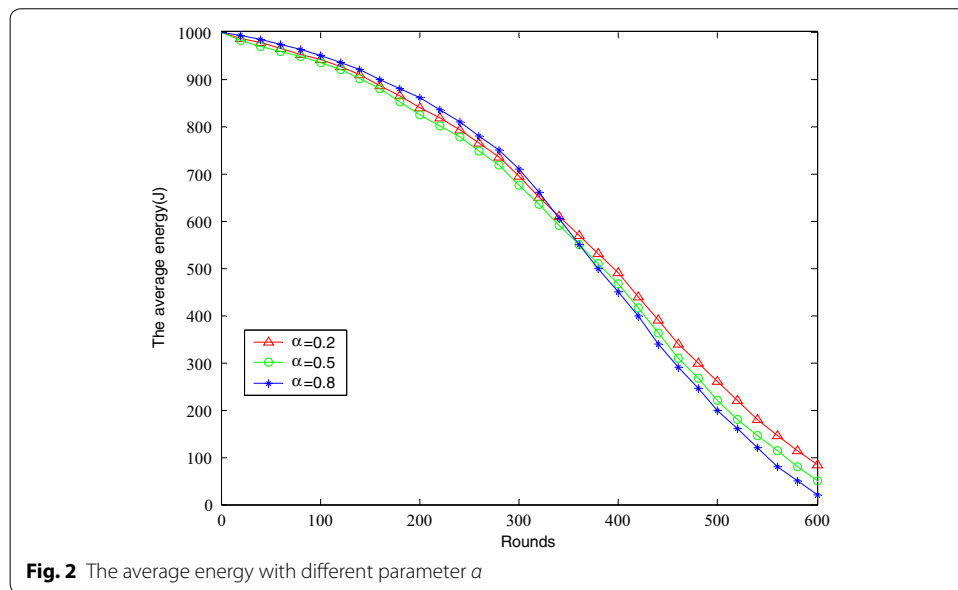
Fig. 2 The average energy with different parameter α

value α is, the greater the proportion of residual energy in the generation of candidate CHs. When the value of α is large, the distance factor becomes more important. For the former case, the nodes near the BS have more residual energy, and the local competition behaves more fiercely, which is bound to generate meaningless energy consumption. For the latter case, the distance factor will lead to the generation of more candidate CHs in the region far away from the BS. In the late stage of network lifetime, the residual energy of nodes typically lies in a low level and is difficult for adaptive reduction of competition radius. In addition, in single hop mode, the nodes far away from the BS will die ahead of time, which is not conducive to the overall energy consumption of the network.
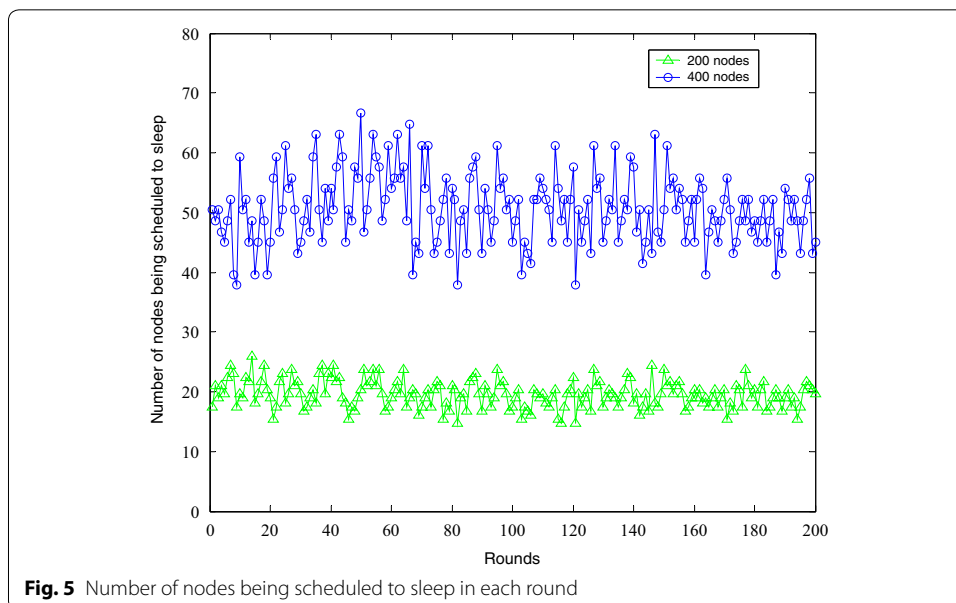
Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 16 of 22

Fig. 3 The variance of sensor's energy with different parameter *a*

Fig. 4 Average transmission radius of the CHs in different regions

According to the previous analysis, select the parameter $\alpha = 0.5$. Under the condition of different node's density, the communication radius of the CH is analyzed, and the results are shown in Fig. 4. It can be seen that the communication radius of CHs is smaller than other CHs near BS. That is because the energy consumption of CH is closely related to the distance between CH and BS in a single-hop manner. The distant nodes from the BS dissipate their energy much quickly due to the long distance transmission. In order to save energy, less CHs should be distributed in the area near BS. However, the measure will lead to the increase of communication distance in that region, which can bring about the burden of the member nodes. On the contrary, in the regions far away from BS, the number of CHs can be promoted to reduce the inter

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18
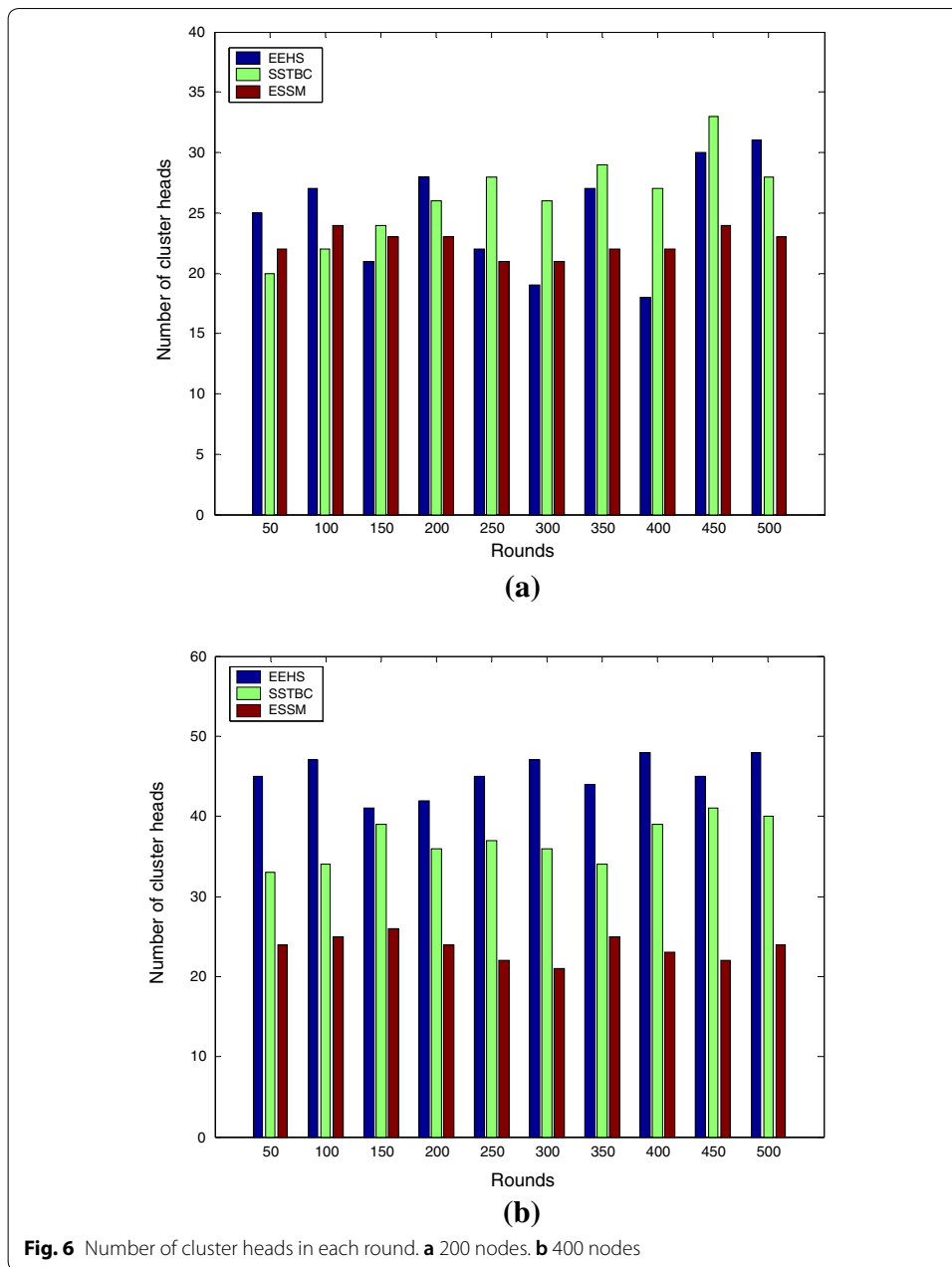
Page 17 of 22

cluster communication range and energy consumption among members. Therefore, the clustering protocol should be designed to compromise the number of CHs in different regions and the communication cost. Our algorithm can make the size of clusters be linearly related to the distance of the BS, which guarantees the little difference of communication radius between the clusters under different node density.

Furthermore, we compare the average number of nodes with different node's density in each round. As shown in Fig. 5, the number of dormant nodes in high-density is more than in low-density. That is because the correlation of data collected by neighboring nodes is much higher and more dormant nodes being selected have little impact on subsequent data fusion. As a result, the number of redundant nodes that are scheduled to sleep is relatively more. In addition, we can find that the average number of dormant nodes at different densities is relatively stable, which provides the basic foundation for maintaining stability of the average energy consumption among all nodes.
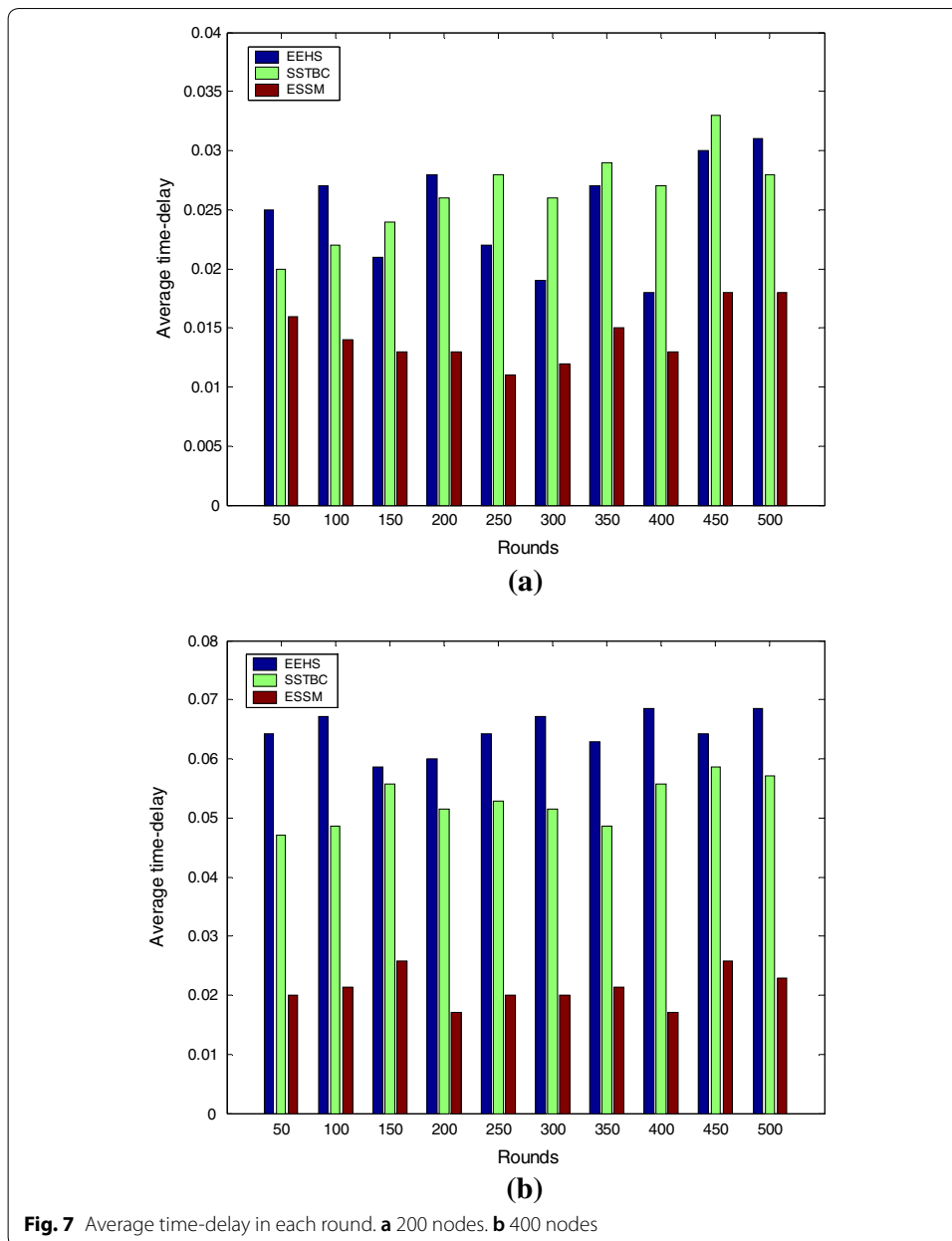
Next, we compare the number of CHs in different node density, and the results are shown in Fig. 6. When the node's density is large, the number of CHs in EEHS is more than SSTBC and ESSM. The reason is that EEHS is mainly concerned with the effect of cluster size on the overall energy consumption. It can be observed that the number of CHs generated in ESSM demonstrates stably during many rounds, and is not affected by the variety of the node density. That is owing to two factors: (1) The CH's selection based on the distance from BS and node's residual energy; (2) When the density is large, more redundant nodes can be scheduled according to the data correlation so as to save energy.

Figure 7 shows that the average time delay. It can be observed that when the node density is large, the time delay of the three algorithms is obviously higher than that of the sparse node's density due to the increase of the network throughput. However, in contrast, the time delay of ESSM can show a distinct advantage. In single-hop mode, CH aggregates the data from its member nodes and transmit it to BS directly, ESSM can reduce delay in aspect of the efficiency of data forwarding.

Fig. 5 Number of nodes being scheduled to sleep in each round

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 18 of 22

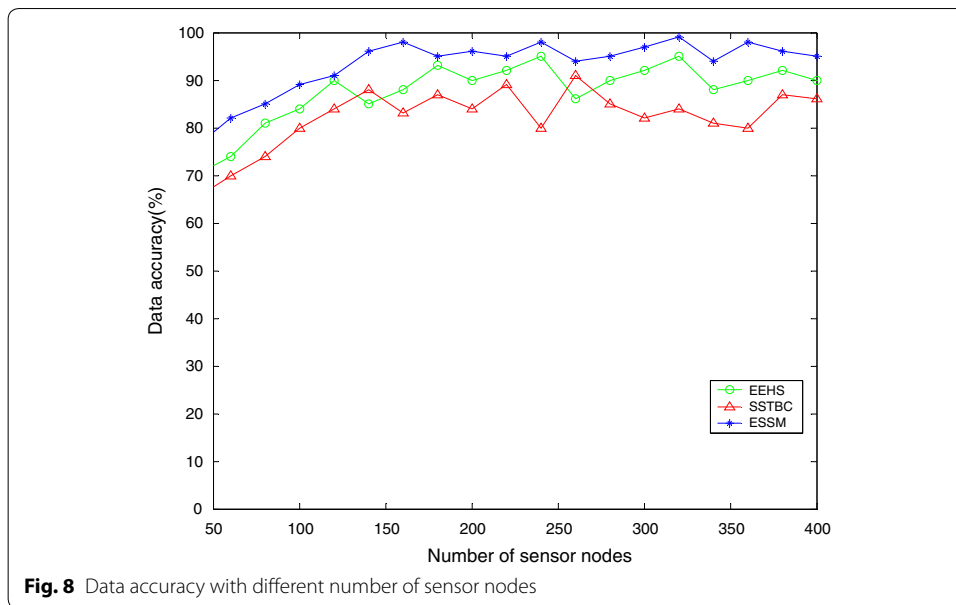**Fig. 6** Number of cluster heads in each round. **a** 200 nodes. **b** 400 nodes

The accuracy metric is defined as the ratio of the result by the data aggregation scheme used to the real summation of all individual sensor nodes in [30, 31]. Figure 8 illustrates the data accuracy of EEHS, SSTBC and ESSM with different number of sensor nodes. From the results, we can observe that the accuracy increases as the number of sensor nodes increases. Obviously, the major reasons is that the data accuracy can reach a certain saturation level till sufficient sensor nodes are deployed for monitoring the triggered event in the sensing field. Besides, ESSM demonstrates better performance in aspect of data accuracy than EEHS and SSTBC. In our proposed scheme, the redundant nodes are selected based on data correlation, and the rest of active nodes are sufficient for achieving the same level of estimated data accuracy. Hence it is unnecessary to make all the

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 19 of 22

Fig. 7 Average time-delay in each round. **a** 200 nodes. **b** 400 nodes

sensor nodes active, which causes an improvement of energy conservation with the prerequisite of data accuracy.

Conclusions and future work

In this paper, we proposed an energy-efficient sleep scheduling mechanism with similarity measure for WSNs. According to the similarity of the data collected by the nodes, the nodes are classified and the redundant nodes can be selected by defining correlation function in fuzzy theory. Therefore, ESSM can activate a minimum number of sensor nodes in a densely deployed environment and maintain high data accuracy. In addition, it improves network lifetime by performing optimal CH selection approach and decentralized sleep scheduling mechanism. However, for the applications with sparse

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 20 of 22

Fig. 8 Data accuracy with different number of sensor nodes

distribution of nodes or low spatio-temporal correlation of data, ESSM will be very hard to evaluate the differences between nodes by their data similarity. And then, the selection of redundant sensor nodes within clusters may not be very effective.

In the future, our work will focus on heterogeneous sensor networks composed of different types of sensors and discuss the synchronization scheme and applies to application scenarios with strict coverage requirements.

Authors' contributions
RW proposes the innovation ideas and theoretical analysis. NX and NTL The carry out experiments and data analysis. All authors read and approved the final manuscript.

Author details
[1] Hubei Co-Innovation Center of Information Technology Service for Elementary Education, Hubei University of Education, Wuhan, China. [2] Department of Mathematics and Computer Science, Northeastern State University, Tahlequah, OK, USA. [3] Faculty of Information Technology, Hanoi National University of Education, Hanoi, Viet Nam.

Competing interests
The authors declare that they have no competing interests.

Ethics approval and consent to participate
Not applicable.

Appendix

Proof for selection of redundant nodes

Lemma *For K categories, a set of redundant nodes $s_*^{(1)}, s_*^{(2)}, \ldots, s_*^{(K)}$ is obtained, then $s_*^{(1)}, s_*^{(2)}, \ldots, s_*^{(K)}$ is contained in the optimal solution Γ.*

Proof For the category k, if the sensor node being selected meets the minimum value of $\sum_{i=1}^{\left|s^{(k)}\right|} Del(s_i^{(k)}, s_*^{(k)})$, the set of redundant nodes can satisfy the maximum amount of

information. Suppose that the proposition is correct for any K categories. Let $s_*^{(1)}, s_*^{(2)}, \ldots, s_*^{(K)}$ denote the redundant set of nodes, which makes the objective function minimum. Then, there exists an optimal solution $\Gamma = \left\{ s_*^{(1)}, s_*^{(2)}, \ldots, s_*^{(K)} \right\} \cup \Psi$. If s' is

refer to as the remaining activity that is compatible with $s_*^{(1)}, s_*^{(2)}, \ldots, s_*^{(K)}$, that is

$$s' = \left\{ j| \sum_{i=1}^{|s^{(v)}|} Del(s_i^{(v)}, s_j^{(v)}) \geq \sum_{i=1}^{|s^{(v)}|} Del(s_i^{(v)}, s_*^{(v)}) \right\} \tag{17}$$

In this way, Ψ is an optimal solution for s'. Otherwise, suppose s' has a solution Ψ', and $|\Psi'| > |\Psi|$. After replacing Ψ with Ψ', the solution $\left\{ s_*^{(1)}, s_*^{(2)}, \ldots, s_*^{(K)} \right\} \cup \Psi'$ will be more

active than Γ. That is in contradiction with the optimal solution of Γ. Thus, Γ is the optimal solution of the selection of redundant nodes.

References
1. Nurelmadina N, Nafea I, Younas M (2016) Evaluation of a channel assignment scheme in mobile network systems. Hum Cent Comput Inf Sci 6(21):1–15
2. Dhasian HR, Balasubramanian P (2013) Survey of data aggregation techniques using soft computing in wireless sensor networks. IET Inf Secur 7:336–342
3. Arunraja M, Malathi V, Sakthivel E (2015) Energy conservation in WSN through multilevel data reduction scheme. Microprocess Microsyst 39:348–357
4. Tyagi S, Kumar N (2013) A systematic review on clustering and routing techniques based upon LEACH protocol for wireless sensor networks. J Netw Comput Appl 36:623–645
5. Vanus J, Belesova J, Martinek R (2017) Monitoring of the daily living activities in smart home care. Hum Cent Comput Inf Sci 7(30):1–30
6. Zheng H, Guo W, Xiong N (2017) A kernel-based compressive sensing approach for mobile data gathering in wireless sensor network systems. IEEE Trans Syst Man Cybern Syst 99:1–13
7. Heinzelman WB, Chandrakasan AP, Balakrishnan H (2002) An application-specific protocol architecture for wireless microsensor networks. IEEE Trans Wireless Commun 1(4):660–670
8. Kumar D, Aseri TC, Patel RB (2009) EEHC: energy efficient heterogeneous clustered scheme for wireless sensor networks. Comput Commun 32(4):662–667
9. Karaca O, Sokullu R, Prasad NR, Prasad R (2012) Application oriented multi criteria optimization in WSNs using on AHP. Wireless Pers Commun 65(3):689–712
10. Hou R, Ren W, Zhang Y (2009) A wireless sensor network clustering algorithm based on energy and distance. In: Second international workshop on computer science and engineering(IWCSE). pp 439–442
11. Wu Y, Fahmy S (2009) Optimal sleep/wake scheduling for time-synchronized sensor networks with QoS guarantees. IEEE/ACM Trans Netw 17(5):1508–1521
12. Wan RZ, Lei JJ, Xu QW (2013) An energy-efficient unequal clustering algorithm in wireless sensor networks. Adv Mater Res 629:801–807
13. Sajjanhar U, Mitra P (2007) Distributive energy efficient adaptive clustering protocol for wireless sensor networks. In: International conference on mobile data management. pp 326–330
14. Ali MS, Dey T, Biswas R (2008) ALEACH: advanced LEACH routing protocol for wireless microsensor networks. In: Asia simulation conference/7th international conference on system simulation and scientific computing. pp 909–914

Wan *et al. Hum. Cent. Comput. Inf. Sci.* (2018) 8:18

Page 22 of 22

15. Lee D, Yoon K (2017) An efficient spatio-temporal index for spatio-temporal query in wireless sensor networks. KSII Trans Internet Inf Syst 11(10):4908–4928
16. Yu T, Wang X, Shami A (2017) Recursive principal component analysis-based data outlier detection and sensor data aggregation in IoT systems. IEEE Internet of Things Journal 4(6):2207–2216
17. Paul S, Sao NK (2011) An energy efficient hybrid node scheduling scheme in cluster based wireless sensor networks. In: World congress on engineering (WCE 2011), vol 2. pp 1775–1779
18. Tan ND, Viet ND (2015) SSTBC: sleep scheduled and tree-based clustering routing protocol for energy-efficient in wireless sensor networks. In: IEEE international conference on computing & communication technologies—research, innovation, and vision for the future (RIVF). pp 180–185
19. Wu X, Cho J, Auriol BJ (2007) Sleep nodes scheduling in cluster-based heterogeneous sensor networks using AHP. In: 4th international conference on embedded software and systems, vol 4523. Springer, Heidelberg, pp 437–444
20. Ha RW, Ho PH, Shen XS (2006) Sleep scheduling for wireless sensor networks via network flow model. Comput Commun 29:2469–2481
21. More A, Raisinghani V (2017) A node failure and battery-aware coverage protocol for wireless sensor networks. Comput Electr Eng 64:200–219
22. More A, Raisinghani V (2014) Random backoff sleep protocol for energy efficient coverage in wireless sensor networks advanced computing. Netw Inf 2:123–131
23. Wu M, Tan L, Xiong N (2016) Data prediction, compression, and recovery in clustered wireless sensor networks for environmental monitoring applications. Inf Sci 329:800–818
24. Shabdanov S, Rosenberg C, Mitran P (2011) Joint routing scheduling, and network coding for wireless multihop networks. In: IEEE Ninth international symposium on modeling and optimization in mobile, ad hoc, and wireless networks (WiOpt). pp 33–40
25. Khalil EA, Ozdemir S (2013) Energy aware evolutionary routing protocol with probabilistic sensing model and wake-up scheduling. In: IEEE Globe-com workshops (GCWkshps). pp 873–878
26. Bagci H, Yazici A (2013) An energy aware fuzzy approach to unequal clustering in wireless sensor networks. Appl Soft Comput 13:1741–1749
27. Imani M, Joudaki M, Arabnia HR (2017) A survey on asynchronous quorum-based power saving protocols in multi-hop networks. J Inf Process Syst 13(6):1436–1458
28. Goyal M, Yadav D, Tripathi A (2017) An intuitionistic fuzzy approach to classify the user based on an assessment of the learner's knowledge level in e-learning decision-making. J Inf Process Syst 13(1):57–67
29. He B, Li Y, Huang H, Tang H (2014) Spatial–temporal compression and recovery in a wireless sensor network in an underground tunnel environment. Knowl Inf Syst 41:449–465
30. Fan Q, Xiong N, Zeitouni K (2016) Game balanced multi-factor multicast routing in sensor grid networks. Inf Sci 367:550–572
31. Cheng H, Su Z, Xiong N, Xiao Y (2016) Energy-efficient nodes scheduling algorithms for wireless sensor networks using Markov random field model. Inf Sci 329:461–477