RESEARCH　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

CrossMark

A new clustering routing method based on PECE for WSN

De-gan Zhang[1,2], Xiang Wang[2], Xiao-dong Song[2], Ting Zhang[3*] and Ya-nan Zhu[2]

Abstract

A new clustering routing method based on predictive energy consumption efficiency (PECE) for a wireless sensor network (WSN) is presented in this paper. It consists of two stages: cluster formation and stable data transfer. In the cluster formation stage, we design an energy-saving clustering routing algorithm based on the node degree, the relative distance between nodes, and the rest energy of nodes. When this algorithm selects the cluster head, the node degree and the relative distance between the nodes are fully considered, so the selected cluster not only has better coverage performance but also short average distance from other member nodes in the formative cluster; therefore, the cost of communications within the clusters is small. In the stable data transfer stage, by using bee colony optimization (BCO), we design a PECE strategy for data transmission. On the basis of considering the predictive values of energy consumption, the hops, and the propagation delay on this route, this strategy gives a precise definition of the route yield by using two types of bee agent to predict the route yield of each routing path from the source node to the sink node. Through the optimization design of the algorithm, it can improve the quality of clusters, thereby increasing the overall network performance, and reduces and balances the energy consumption of whole network and prolongs the survival time of the network.

Keywords: WSN; BCO; PECE; Clustering routing

1 Introduction

A wireless sensor network (WSN) is composed of a large number of low-cost and low-power wireless sensor nodes. They are deployed in the specified area and form a wireless network by way of self-organizing. They can work normally at a wicked or special environment that people cannot close. This technology has been widely used in the industry, military, environmental monitoring, and medical and other fields. These nodes can be deployed easily, but they can only use the battery, and it is difficult to change the battery, so how to prolong the life cycle of the whole network is one of the hot research topics in WSN [1, 2].

WSN routing protocol is responsible for looking for a data transfer path in the network layer. The data packet from the source node is forwarded to the data-receiving node by way of multi-hop communication on this path. The stand or fall of routing protocol directly determines

the value of energy consumption when the sensor nodes are transmitting data and the survival time of the network [3]. For these characteristics of the nodes in network, like random deployment, limited energy, self-organizing, and frequent changes in network topology, there are better adaptability and energy efficiency by using a hierarchy routing algorithm based on clustering than using the plane routing algorithm [4]. Clustering algorithm is to divide the sensor network nodes into different clusters. Every cluster has a cluster head node, and the other member nodes send information to the cluster head node which continues to fuse and forward these data. Among them, it is the key of clustering algorithm to select a cluster head, and research about how to lower the node energy consumption by selecting the cluster head and then form a high-quality cluster has an important significance.

Swarm intelligence (SI) is a kind of compute intelligent algorithm based on the research of the collective behavior of social insects in a decentralized and self-organizing system, as described in [5]. The ant colony optimization [6] and the bee colony optimization [7]

* Correspondence: 172128173@qq.com
[3]Department of Computer Science and Technology, Tangshan College, Tangshan 063200, China
Full list of author information is available at the end of the article

Zhang *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:162

Page 2 of 13

have been widely used in network colony intelligence technology.

Swarm intelligence is used to solve a given problem by a kind of collective act of autonomous agents which interact with each other in the distributed environment, in order to find a global solution to this problem, as defined in [8]. The design algorithm of swarm intelligence is inspired by the collective act of the insects (such as bees, termites, and ants) which exist in a decentralized and self-organizing system or other animal society [9–15]. These insects live in a hostile and dynamic environment; they survive through coordination and cooperation. They communicate with each other directly or indirectly via the environment to communicate in order to complete their basic tasks, such as foraging, division of labor, nesting, or hatching classification [15–20].

The bee colony optimization (BCO) model is a new paradigm of swarm intelligence. It mainly requires two types of agents that are used for routing [21–26]: the scouts find out the new route to sink node as required, and the foragers will transmit the data packets and evaluate the quality of the route that has been found by the predictive energy consumption of the routing path and the end-to-end delay [27–36]. The foragers perceive the network state and evaluate the different routes in WSN according to the measured indicators, and then choose the suitable path for routing data packets with the intent to maximize the survival time of the network [37–45].

We propose a novel clustering routing approach based on predictive energy consumption efficiency (PECE) for a wireless sensor network in this paper. First, in the cluster formation stage, from the perspective of how to choose a cluster head, we design an energy-saving clustering routing algorithm based on the node degree, the relative distance between nodes, and the rest energy of nodes. This algorithm ensures the uniform distribution of cluster heads and the balance of cluster size, and the high-energy nodes have priority to be the cluster head. Then, in the stable data transfer stage, inspired by the process that the bees look for their food, we design a new strategy for predictive energy-efficient data transmission, in order to find all the possible paths from a source node to a sink node and select the optimal path from them. This strategy of data transmission is based on two basic parameters: the energy consumption of the routing path and the end-to-end delay. These two parameters reflect the yield of the path that has been distributed to the forager agent.

2 Related works
2.1 LEACH method for clustering routing
In the current typical algorithms for clustering routing, the LEACH (Low-Energy Adaptive Clustering Hierarchy) proposed by [9] is one of the most representative clustering algorithms. This is an adaptive topology algorithm; it executes the process of cluster reconstruction periodically. In the process of cluster formation, the nodes in the WSN have the same probability to act as the cluster head. The selection of the cluster head is random, so the energy consumption of nodes is balanced. However, the random selection mechanism does not consider the rest energy of nodes and cannot guarantee the scalar rationality and the distributional uniformity of the cluster head node [10]. On this basis, some researchers propose the LEACH-ED [11] protocol based on the distance from the source node to the cluster head node and the rest energy, and the LEACH-T [12] protocol which uses the interval instead of the random number of cluster head and the threshold. But the two improved protocols also have not considered the distributive rationality of the cluster head node.

2.2 PEGASIS protocol
The PEGASIS (Power-Efficient GAthering in Sensor Information Systems) protocol [13] proposed by Lindsey etc. makes a chain which has a minimum sum of the length with all the nodes in WSN by a greedy algorithm. Each node of the chain just sends and receives data only once and sends data with a minimum power. The nodes select a cluster randomly to communicate with the base station each round, reducing the data traffic. Experiments proved that PEGASIS has a nearly double life cycle than LEACH. But it has a routing cavity problem just like that shown in Fig. 1.

2.3 TEEN and EEUC methods
TEEN (Threshold Sensitive Energy Efficient Sensor Network Protocol) [14] adopted a sink searching and message forecast path as shown in Fig. 2. It should set hard and soft thresholds, so it is very complex. EEUC (Energy-Efficient Unequal Clustering) [14] adopted a multi-hop wireless communication method. The energy consumption included inter-cluster and intra-cluster.
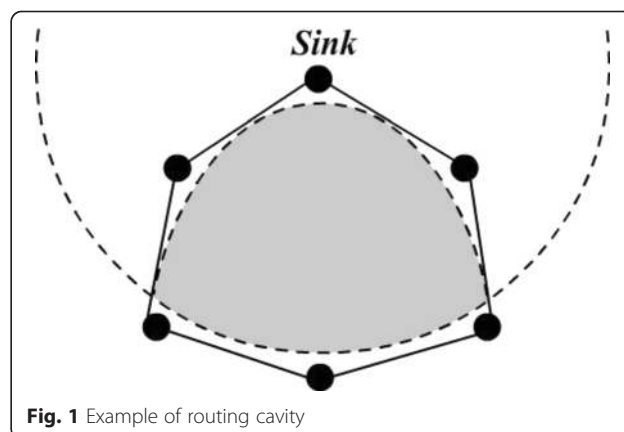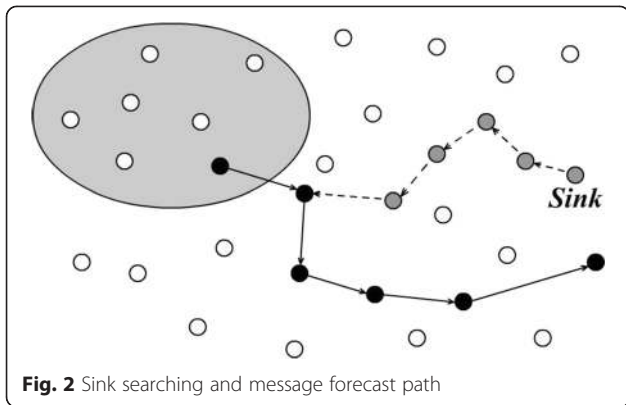

Fig. 1 Example of routing cavity

Zhang *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:162

Page 3 of 13

Fig. 2 Sink searching and message forecast path

When the number of hops is larger, the energy consumption also is larger.
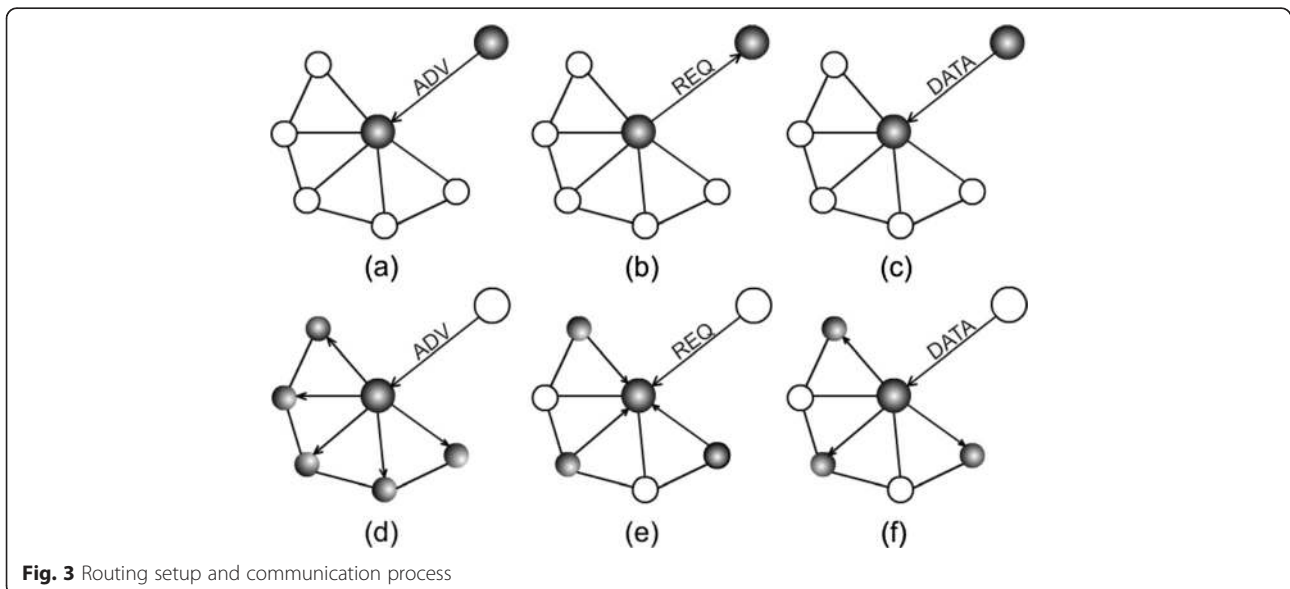
2.4 Other existing schemes

Many previous research efforts have tried to achieve trade-offs in terms of delay, energy cost, and load balancing for such data collection tasks [15–20]. Many recent research efforts on open vehicle routing (OVR) problems [21–26] have been studied, which are active areas in operations research and based on similar assumptions and constraints compared to sensor networks. Several new insights motivate the scholars [27–37] to adapt these schemes so that they can solve or prove certain challenging problems in WSN applications, such as the data collection protocol called EDAL, which stands for Energy-efficient Delay-aware Lifetime-balancing data collection [38–45]. For example, the authors in [38] propose an online, multi-objective optimization (MO) algorithm to efficiently schedule the nodes of a WSN and to achieve maximum lifetime. Instead of dealing

with traditional grid or uniform coverage, they focus on the differentiated or probabilistic coverage where different regions require different levels of sensing. The MO scheme helps to attain a better trade-off among energy consumption, lifetime, and coverage. The scheme can be run every time a node failure occurs due to power failure of the node battery so that it may reschedule the network. This is a good scheme. Due to the length limit of the paper, we cancel comments of the other papers.

3 Principle and model of the approach

By analyzing the shortcomings of typical protocols for clustering routing in WSN mentioned above, combined with the BCO model, this paper proposes a new method, PECE. Compared with the typical protocols, the new protocol improves the quality of clusters, reduces and balances the whole network energy consumption, and prolongs the survival time of the network. Our PECE-based clustering routing approach will adopt routing setup and communication process (including step (a) to step (f)) as illustrated in Fig. 3.

The BCO model is a new generic swamp intelligence (SI) optimization technology, which achieves the effective labor employment and energy consumption through a distributed multi-agent model. The ant colony optimization (ACO) model mainly adopts a natural insectival behavior that is "looking for food" to find the shortest path between the colony and the source of food. Different from the ACO, the BCO mainly adopts two natural behaviors from the social life of bees: the behavior in the process of mating and the behavior in the process of foraging. Mating behavior is actually used as a powerful SI optimization clustering technology, and this technology could compete with other typical algorithms for clustering and other SI optimization

Fig. 3 Routing setup and communication process

Zhang *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:162

Page 4 of 13

models particularly the ACO. The foraging behaviors used in this study are based on the behavior of bees in nature to find the food source. This behavior aims to find the highest quality food source. Inside the hive, the bees are divided into five kinds plus the queen. The bees living in the hive are "food packager" and "caregivers"; both are responsible for feeding the queen and bee brood larvae. The other three kinds of bees take part in the process of searching food: "scouts", "foragers", and "worker bees". Our proposed routing approach will take advantage of the process of foraging bees. There are mainly three types of bees participating in this process:

3.1 Scouts
Scouts are responsible for discovering all possible food sources (all paths). Then, they direct foragers to them from the hive through the "swing dance"; this swing dance indicates the orientation of food.

3.2 Spectator (or forager)
They are responsible for evaluating the discovery food sources (based on nectar quantity and quality) and employ the worker bees and direct them to the position where they are located.

3.3 Worker bees (or hire bee)
Worker bees gather nectar at the evaluated food source following the forager. After giving up a food source, a worker bee can transform into a scout bee and find the next potential food sources.

Inspired by foraging behavior and based on the artificial bee colony (ABC) algorithm [15], Karaboga defined clustering as a process of identifying natural grouping or swarm in multi-dimensional data. Distance measurement is commonly used to assess the similarity between patterns. Given N objects, each object is assigned to one of the K clusters. The sum of squares of the Euclidean distance from each object to their cluster's center can be calculated. The objective of clustering would be to minimize Eq. (1):

$$J(w, z) = \sum_{i=1}^{N} \sum_{j=1}^{K} w_{ij} \| X_i - Z_j \|^2 \qquad (1)$$

In the formula above, W_{ij} is the correlation weight of mode X_i and cluster j, $X_i(i = 1, \cdots, N)$ is the position of the ith mode, and Z_j ($j = 1, \cdots, K$) is the center of the jth cluster, which can be found in Eq. (2):

$$Z_j = \frac{1}{N_j} \sum_{i=1}^{N} w_{ij} x_i \qquad (2)$$

N_j is the number of modes in the jth cluster, and W_{ij} is the correlation weight of mode X_i and cluster j. Its value

is 1 or 0 (if mode i is assigned to cluster j, W_{ij} is 1; otherwise, it is 0).

By minimizing (optimal) the sum of the Euclidean distance between generally instances X_j and the center of cluster Z_j in an N-dimensional space and adjusting to proceed, each solution z_i is a D-dimensional vector; here $i = 1, 2, \cdots,$ SN. D is the number of products and includes the size of the input and the cluster for each data set.

After initialization, the position (solution) group proceeds repeated cycles. An employed bee generates the modification of location (solution) in its memory according to local information (visual information) and tests the nectar volume (fitness value) of a new food source (new solution). If the new food source has more nectar volume than the previous one, the bee will remember the new location and forget the old location information. Otherwise, it retains the location information. Equation (3) gives the cost function of pattern i (f_i):

$$f_i = \frac{1}{D_{\text{Train}}} \sum_{j=1}^{D_{\text{Train}}} d\left(x_j, P_i^{CL_{\text{Known}}(x_j)} \right) \qquad (3)$$

Here, D_{Train} is the number of training patterns, and the training mode is used for normalizing the sum, so that any distances are limited to the range [0.0, 1.0]. $\left(P_i^{CL_{\text{Known}}(x_j)} \right)$ defines the instance which belongs according to the database. The location of a food source represents a possible optimization solution, and the nectar volume of a food source corresponds to the quality (fitness) of relevant solution, calculated by Eq. (4):

$$\text{fit}_i = \frac{1}{1 + f_i} \qquad (4)$$

An artificial spectator bee selects a food source according to the associated probability values. These probability values are expressed in P_i, calculated by Eq. (5):

$$P_i = \frac{\text{fit}_i}{\sum_{n=1}^{\text{SN}} \text{fit}_n} \qquad (5)$$

Here, the SN is the number of food source, and the employed bees have the same number. fit_i is the fitness of solution given by Eq. (4); it is inversely proportional to the f_i given by Eq. (3). In order to select a candidate from old food location information in memory, the artificial bee colony algorithm uses Eq. (6):

$$V_{ij} = Z_{ij} + \phi_{ij}(Z_{ij} - Z_{kj}) \qquad (6)$$

Here, $k \in \{1, 2, \cdots, \text{SN}\}$ and $j \in \{1, 2, \cdots, \text{D}\}$ are the randomly selected targets, D is the number of products,

Zhang *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:162

Page 5 of 13

and the following is the same. Although k is a randomly determined count parameter, it must be different from the count parameter i. ϕ_{ij} is a random number in $[-1, 1]$; it controls the production of the neighbor food sources around Z_{ij} and represents a comparison of two food locations that are visible to a bee. It can be seen from Eq. (6) that when the difference between Z_{ij} and Z_{kj} is reduced, the disturbance of location Z_{ij} also decreases. Therefore, in the search space, when the search approximates optimal solution, the step size adaptively reduces. The food resources which are discarded would be replaced with a new food resource by scouts.

The artificial bee colony algorithm could randomly generate a location to replace the discarded location. This process could be simulated. In this algorithm, if a position cannot be further improved through a predetermined number of cycles, then the food source is assumed to be discarded. The value of a predetermined number of cycles is an important control parameter in this algorithm, called as a "limit" for a discard act.

The source Z_i is assumed to have been discarded and $j \in \{1, 2, \cdots, D\}$, so a new food resource found by scouts replaces the source Z_i, as shown in Eq. (7). This operation can be defined as

$$Z_i^j = Z_{\min}^i + \text{rand}(0, 1)\left(Z_{\max}^j - Z_{\min}^j\right) \tag{7}$$

Each candidate source location V_{ij} comes into being and then is evaluated by artificial bees. Its performance will be compared with the performance of the old location.

4 Design of the algorithm

The network model is adopted as follows:

1) We assume all nodes have the same primary energy and do not move after deployment.
2) Sensor nodes are randomly distributed, and each node has a unique ID code in the whole network.
3) The location of all nodes in the network is unknown, and there is no need to use the positioning system or positioning algorithm to learn the location of nodes.
4) The transmit power of the node is settled, and the approximate distance between one node and the sending node can be calculated according to the received signal strength indication (RSSI).

4.1 The node energy consumption model

The algorithm is based on the wireless communication energy consumption model proposed in the literature [2]. The calculated formulas of energy consumption are as follows:

The energy consumption of sending data

$$E_t(k, d) = \begin{cases} E_{elec} \cdot k + \varepsilon_{fs} \cdot k \cdot d^2, d < d_0 \\ E_{elec} \cdot k + \varepsilon_{mp} \cdot k \cdot d^2, d \geq d_0 \end{cases} \tag{8}$$

$$d_0 = \sqrt{\varepsilon_{fs}^2 \Big/ \varepsilon_{mp}} \tag{9}$$

The energy consumption of receiving data

$$E_r(k) = E_{\text{elec}} \cdot k \tag{10}$$

When the node detects the exchange packets within its scope in monitor mode, its energy consumption is

$$E_0(k) = E_{\text{elec}} \cdot k \tag{11}$$

In the formulas above, k is the number of bytes of the transport packet and d is the distance of transmission. When the transmission distance is less than the threshold value d_0, the power amplification uses the free space propagation model; otherwise, the multi-path fading model is used. E_{elec} (in J/bit) is the factor of RF energy consumption. The ε_{fs} and ε_{mp} are respectively the energy consumption factor of the amplifier circuit in two models.

The distance d from this node to the sending node can be calculated by the RSSI. The formulas are as follows [17, 18]:

$$d = 10^{\frac{|\text{RSSI}-A|}{10 \times n}} \tag{12}$$

In Eq. (12), A is the received signal strength at the position where is 1 m away from the sending point. RSSI is the received signal strength indication, and n is the factor of path fading, which is generally 2 to 5.

Thus, the sum of energy consumption of a node n_i could be calculated by Eq. (13):

$$E(n_i) = E_t(k_{n_i}, d) + E_t(k_{n_i}) + E_0(k_{n_i}) \tag{13}$$

Since the energies which are consumed during the transmission or reception are concerned with the communication process, so they can be considered the efficient energy consumption. However, when detecting the exchanging packets or in idle mode, the energy consumption is invalid to deplete the batteries of nodes. Therefore, this energy consumption should be minimized, and that can be achieved by determining a small "idle" time. After the idle time, the node automatically enters a "sleep" mode to save its battery. In addition, when a WSN increases the number of nodes, the number of neighbors of nodes also increases, which leads to consumption of more energy at detecting or forwarding.

4.2 Design of the algorithm

Like other clustering algorithms, the algorithm takes rounds to work. Each round is divided into two parts: cluster formation and stable data transfer. We introduce

Zhang *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:162

Page 6 of 13

several definitions used in this study. M is the set of all nodes in the network.

Definition 1 Neighbor nodes: In M, $d_{(i, j)}$ represents the Euclidean distance between any two nodes i and j. When $d_{(i, j)} \leq R_c$, the node j is called the neighbor node of node i. Here, R_c is the broadcast radius of the node and the $Neighbor_i$ represents a set of neighbor nodes of node i.

Definition 2 Node degree: In M, the number of nodes which are contained in the R_c range of any node i is called the node degree, expressed with $Number_i$. The higher the $Number_i$ is, which represents that there are more nodes surrounding node i, the better coverage performance the cluster whose head node is i has.

Definition 3 The relative distance between nodes: In M, the node i receives all the broadcast messages sent by nodes which are in the R_c range of node i, and uses SS_j to represent the signal strength from node j. I_i is defined as follows:

$$I_i = \sum_{j=\text{Neighbor}_i} SS_j / \text{Number}_i$$

The higher the I_i is, which means that the average distance between node i and its surrounding nodes is shorter, the less the energy consumption is when they are communicating.

Definition 4 Member nodes: After the process of cluster head selection, if node j is in the coverage area of cluster head node i, in other words, the distance between j and i is less than R_c, so the node j is called the member node of node i, provided that any node can become a member node of only one cluster.

4.2.1 The stage of cluster formation

In the initial stage, assuming that all nodes in the network have the same clock, specific implementation steps are as follows:

Step 1 Information broadcasting stage: All the nodes broadcast its own ID information to the outside. In this process, the node i will count its $Neighbor_i$, $Number_i$, and I_i by received information.

Step 2 Role determining stage: All nodes execute the following operations:
  1) Each node calculates its own regular time t according to Eq. (14) and start timing after

receiving the information sent from the base station. If the node i does not receive the information sent from cluster head within the time t_i, then it declares itself to be the cluster head and informs its neighbor nodes.
  2) If it receives information sent from the cluster head, so this node selects to be the member node and exits timing.
  3) If it receives the information sent from multiple cluster heads, then it chooses to join the cluster which sends the information to this node last.

In the above, t_i is calculated by the following formula:

$$t_i = \alpha \times e^{-w_i} \tag{14}$$

In this formula, α is the scale factor that determines the size of delay, wherein w_i is as follows:

$$
\begin{aligned}
w_i = 100 \\
\times \left(C_1 \cdot 1/d(i) + C_2 \cdot (1 - 1/\text{Number}_i) + C_3 \cdot e^{E_i} \right)
\end{aligned}
\tag{15}
$$

In the above formula, the E_i represents the current remaining energy of node i; $C_1 + C_2 + C_3 = 1$, $d(i) = 10^{|\text{RSSI}_i - A|/(10 \times n)}$.

Equations (14) and (15) show that the nodes with higher degree, shorter average distance away from surrounding nodes, and more rest energy have shorter waiting time t and more chance to be the cluster head; thus, that ensures the rationality of selecting the cluster head.

When a member node receives the information sent from multiple cluster heads, then it chooses to join the cluster which sends the information to this node last, because the approach for cluster head selection in this paper makes the size of the preferred cluster less than the later-formed cluster, so choosing to join the later-formed cluster can balance the size of each cluster.

4.2.2 Stable data transfer stage

After cluster building is completed, the cluster head node creates the TDMA schedules based on the number of member nodes and informs member nodes of the time slot of sending data. The member nodes send data only in the allocated time slot, and then during the rest of the time, they are in a dormant state to conserve energy, and the data sent by member nodes are integrated at the cluster head node and sent to the sink node finally by the cluster head node.

The path discovery process uses a bee agent. The process that one source node sends the data packet to the sink node may exhaust the energy of all the nodes along the path. If the optimal path from a source node to the sink node only depends on the number of hops without considering the energy of the node battery,

which may cause the data packet loss in the transmission, this leads to a longer delay to re-route data packets. In addition, when the overloaded nodes stop running, the unfair distribution of network traffic will lead to network partitioning.

Therefore, a fault-tolerant and efficient routing protocol should consider the other route's information of energy consumption before choosing a path to transmit data. In order to hire the artificial bee agent from the source nodes to the sink node, this study uses the BCO model. The bee agents travel on all the possible paths to collect energy information of all the nodes along the path, predicting the routing energy consumption and selecting the optimal path. The energy information of a path should be displayed:

1) Remaining battery power of each node: If it is less than a predetermined threshold value, then the path cannot be selected to transmit packets.
2) The total energy consumption of path nodes: In order to route packets on the path which consumes less energy, this parameter will show the efficiency of the path from the angle of energy. The path which consumes less energy tends to have the minimum number of hops, because it will go through the least number of nodes.

Selecting the path which consumes less energy and thus saving battery of nodes along the path prolong the network lifetime by extending the battery life of nodes.

In Fig. 4, there are three possible paths from the source node "A" to the sink node "S": path R_1: A, B, E, S; path R_2: A, C, F, S; and path R_3: A, D, G, I, S.

The selection of optimal path in these three paths depends on the predictive energy consumption on this path in the process of communication. This energy information will be collected by the foregone bee agent at the journey of path founding, and it mainly include two essential metrics: the rest energy of nodes, which is used to determine the predictive lifetime of nodes and thus determine the efficiency of nodes in the process of transferring packets on the path, and the predictive total energy consumption of path nodes. For each path, it is expressed in $E(R_j)$, where j is the index of the path.

In order to calculate the total energy consumption of all the nodes along each path in the process of receiving, Eq. (16) has given the calculation method, and the formula can be applied to the three possible paths in the previous example: R_1, R_2, and R_3.

$$E(R_j^j) = h(R_j^j) \times E_r(k) \qquad (16)$$

In this formula, $h(R_j)$ is the hops on the path R_j and $E_r(k)$ is the energy consumption in the process of receiving packets whose size is k bytes.

If $E(R_1) < E(R_2) < E(R_3)$, then R_1 is the routing path with the lowest energy consumption. If it meets all constraints of other threshold values, for example, in order to achieve reliable packet transmission along the path in the communication process, the node remaining battery power P(n) and the minimum propagation delay $D(R_j)$, then R_1: A, B, E, S will be selected as the optimal path from source node A to sink node S. So the path yield $g(R_j)$ can be a rate of each potential path assigned from A to S and combined: the predictive total energy consumption of path nodes, the hop count, and the
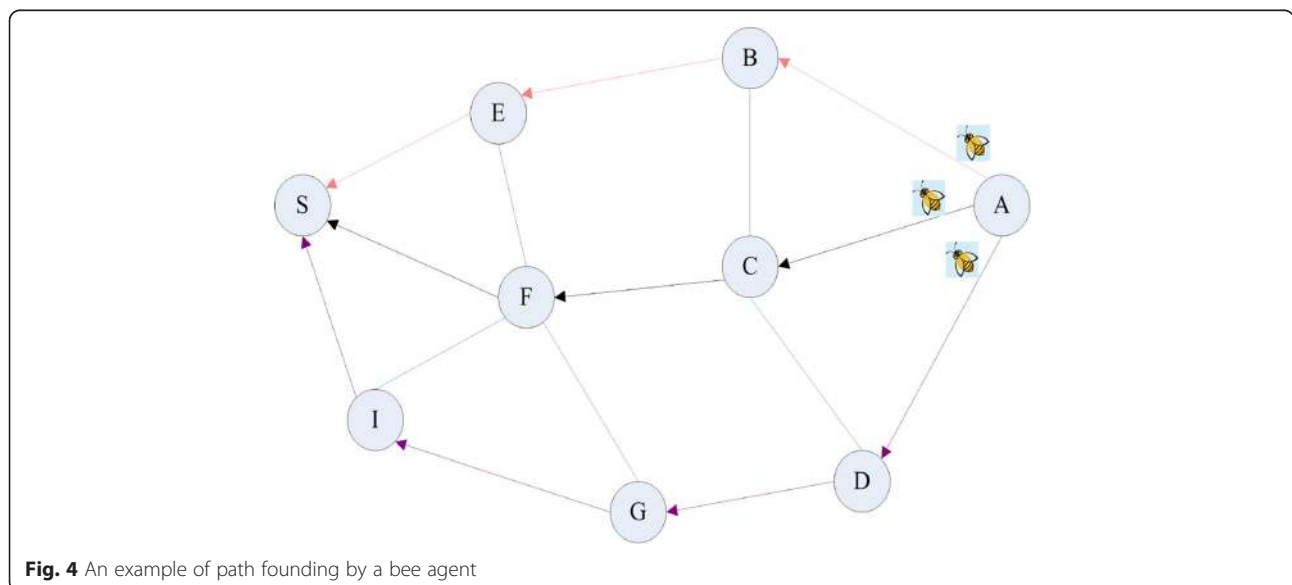

Fig. 4 An example of path founding by a bee agent

propagation delay. The end-to-end propagation delay can be represented and calculated by Eq. (17).

$$D(R_j) = \sum_{i=1}^{N} d(n_{ji}, n_{ji+1}) \tag{17}$$

In the formula, N is the number of nodes on the path R_j and $d(n_{ji}, n_{ji+1})$ is the propagation delay from node i to node $i+1$ on the same path j. Finally, the yield $g(R_j)$ of path R_j in the M paths can be represented by Eq. (18):

$$g(R_j) = \frac{E(R_j) \times D(R_j)}{\sum_{j=1}^{M} E(R_j) \times D(R_j)} \tag{18}$$

Therefore, for each potential path from the source node to the sink node, in order to reflect its energy consumption and propagation delay, each potential path should be given a yield. The optimal path R_0 is the path with the highest yield, given as Eq. (19):

$$R_0 = \max(g(R_j)) \tag{19}$$

It is worth mentioning that the queuing delay and transmission delay can be represented by a constant. This is because the process is in the route discovery stage, rather than the stage of routing data on the selected path, that does not refer to any delay caused by queuing and transmission.

Inspired by the SI model and the BCO model, the predictive energy-efficient model proposed by this paper is an optimal model-oriented WSN. It can be summarized as shown in Fig. 5. In the flow diagram, the process of optimal path discovery from the source node n_s to the sink node S is as follows:

1) For each source node n_s, in order to route its data packets to the sink node S efficiently, it can start the process of routing path discovery to select the optimal path in all possible M paths which can reach to the sink node S.

2) For the path discovery, each node sends the bee agent (via beacon message) associated with a TTL (time to live, predefined to prevent long delays and increased routing overhead) to all the neighbor nodes in the M potential paths. Bee agents (scouts) will collect and store all the required routing information, including the remaining battery power which is considered a key indicator, and other information, such as queuing delay.

3) If the TTL data packet expires, the data packet of the bee agent will indicate the failure to the source node and the path is denied.

4) When a bee agent reaches to the sink node S, after collecting all the necessary routing information, it is sent to its source node n_s by the same recorded path. And then the bee agent (becoming a forager) will reveal the path information founded by itself according to the nodes of every potential path: the remaining battery power $P(n_i)$, the hops $h(P_j)$, and the end-to-end delay $D(P_j)$ in each path j.

5) At the source node, the amount of energy that will be consumed should be calculated as a function, the energy consumption of this function is $h(P_j)$ (the hops) times the amount of the energy consumption of each node on the path given by Eq. (13).

6) Energy consumption $E(R_j)$ will be calculated by Eq. (16), and the propagation delay $D(R_j)$ is calculated using Eq. (17).

7) Finally, the yield $g(R_j)$ of each path will be inferred by foragers according to Eq. (18) to determine the optimal path R_0. The optimal path is the path with the highest yield, and the highest yield is given by Eq. (19) based on the predictive energy consumption, the hops, and end-to-end delay, so as to route data packets from the source node n_s to the sink node S on this path. During transmission, if there are any problems or faults occurring on the optimal path, other potential routing paths can also be adopted, but this is related to their yields.

5 Test and analysis of the results

5.1 Settings of simulation parameters

In order to verify the performance of the algorithm, the technical equipment used by us are Crossbow, MicaZ, and Iris, and the simulators used by us are TOSSIM, OMNeT++, and ATEMU. These technical equipment and the simulators are often used by our research group. Here, we show the simulations by OMNeT++. We distribute 100 wireless sensor nodes randomly in a 100 m × 100 m area. The sink node is set on the center of the area [19]. The simulation parameters are shown in Table 1. Here, we only show the analysis of the relationship between the number of clusters and the survival time, the analysis of the energy efficiency and the survival time, and the comparative analysis of energy consumption balance. The other more considered metrics or factors have been canceled because of the length limit of the paper.

5.2 Analysis of the relationship between the number of clusters and the survival time

According to our analysis, we can obtain the optimal value $R_c \in [16, 54]$ of the broadcast radius of the sensor node. Here, we provide the $R_c = 30$ m to simulate. According to our analysis, in the case of $R_c = 30$ m, the ideal number of optimal clusters is $n = 6$. The test (Fig. 6)

Fig. 5 Strategy for data transmission based on the BCO model
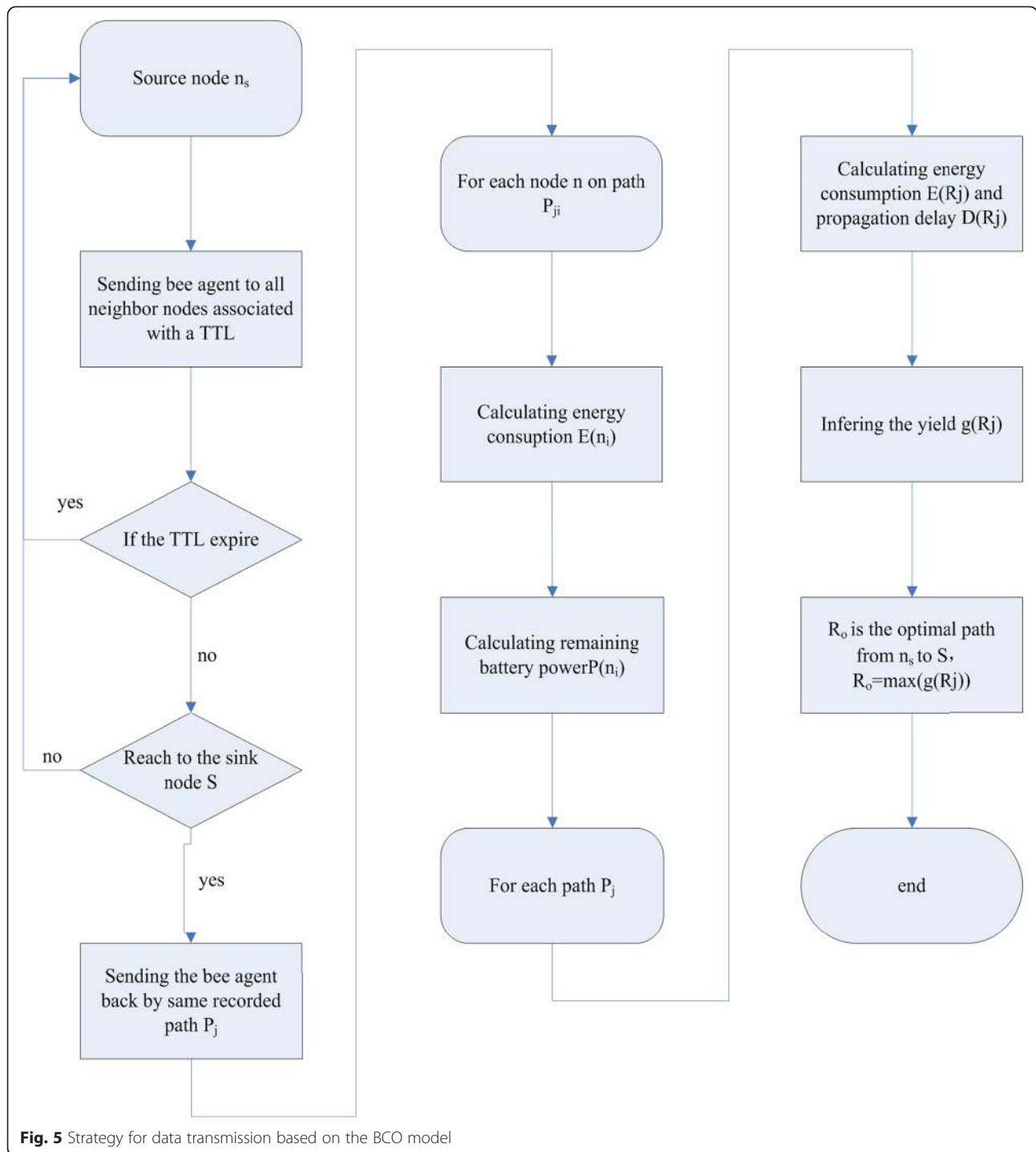
of relationship between the number of clusters and the survival time shows the following: with the increase of the number of clusters, the survival time of the network has a rapid promotion, and when the number of clusters is 6, the survival time of the network reaches up to a peak of 1732, which is consistent with theoretical derivation and thus verifies the correctness of the algorithm in this paper.

Survival time begins to decline when the number of clusters increases to 75. The survival time starts to be smooth and is maintained at about 150 rounds. The reason for such result generation is as follows: when the number of clusters is less than 6, although the communication between clusters and the number of nodes on the communication between clusters are both reduced, the communication radius between neighboring cluster

Zhang *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:162

Page 10 of 13

Table 1 Experimental parameters

Parameter	Value
Packet length	2000 bit
Node's initial energy E_0	0.5 J
Node's distribution range	100 m × 100 m
Sink node coordinate	(50, 50)
Node number	100
The factor of RF energy consumption E_{elec}	5.0×10^{-8} J/bit
Data fusion energy consumption E_{DA}	5.0×10^{-9} J/bit
The parameter of channel propagation model	ε_{fs}, 1.0×10^{-11} J/(bit/m^2); ε_{fs}, 1.0×10^{-11} J/(bit/m^2)

heads becomes longer, which increases the energy consumption of communication between clusters. However, when the number of clusters is more than 6, despite the communication radius between neighboring clusters becoming smaller, the communication between clusters and the number of nodes on the communication between clusters are both increased.

When the number of clusters of the entire network is 6, the chain length in clusters accounts for about 17 % of all the number of nodes, the optimal performance achieves. When the number of clusters in the network reaches to 75, the average number of nodes within a cluster is 1 to 2, and the survival time of the network approaches the conditions which are under the algorithm of flat routing.

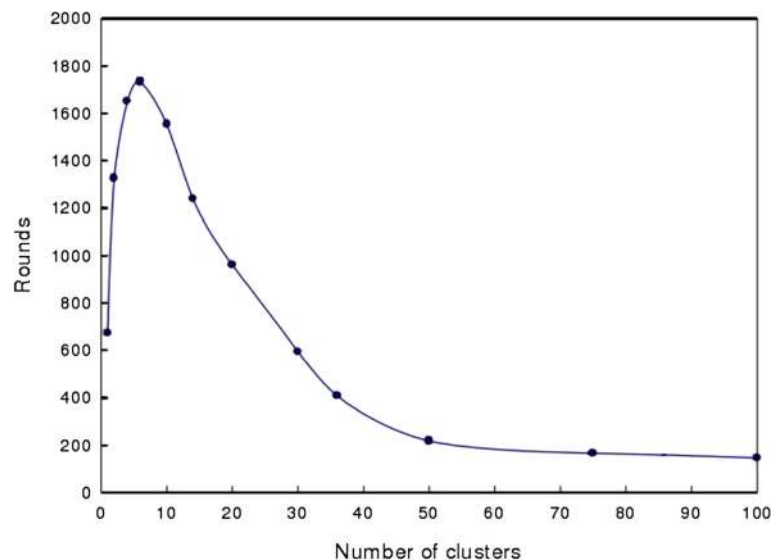### 5.3 Analysis of the energy efficiency and the survival time

The network lifetime and the balance of energy consumption of nodes reflect the overall performance of the network. In the test of network performance comparison between PECE (shown as PEECR in Figs. 7 and 8), LEACH, and PEGASIS, we provide the number of clusters as 6. As shown in Fig. 7, the dead nodes generate respectively at the 343rd round with LEACH and at the 679th round with PEGASIS. However, the first dead node generates at the 1589th round with PECE, and it extends to the 1246th round with LEACH and at the 910th round with PEGASIS. The network lifetime increases substantially. All the nodes are dead when the LEACH and PEGASIS networks run respectively to the 574th round and the 1541st round. The rounds from the first dead node to the last dead node are respectively 233 and 862. All the nodes are dead in the PECE network at the 1732nd round. The process of node death is experienced at the 143rd round. It can be seen that the process of node death with PECE experiences much less time than LEACH and PEGASIS; the network energy decreases evenly.

In addition, using the PECE algorithm (Fig. 8), the energies of the network ran out at about the 1700th round, and the PECE algorithm could keep more time and more balance of energy consumption than LEACH and PEGASIS; the nodes are more energy-efficient.

5.4 Comparative analysis of energy consumption balance

Based on our experiments, we know that using the LEACH protocol in the network, the forgoing 683 rounds is the network lifetime before the dead nodes have generated. Although the energy consumption of each round is fluctuating around the average of 0.23 J, the D value of energy consumption between different rounds is very large: the highest energy consumption

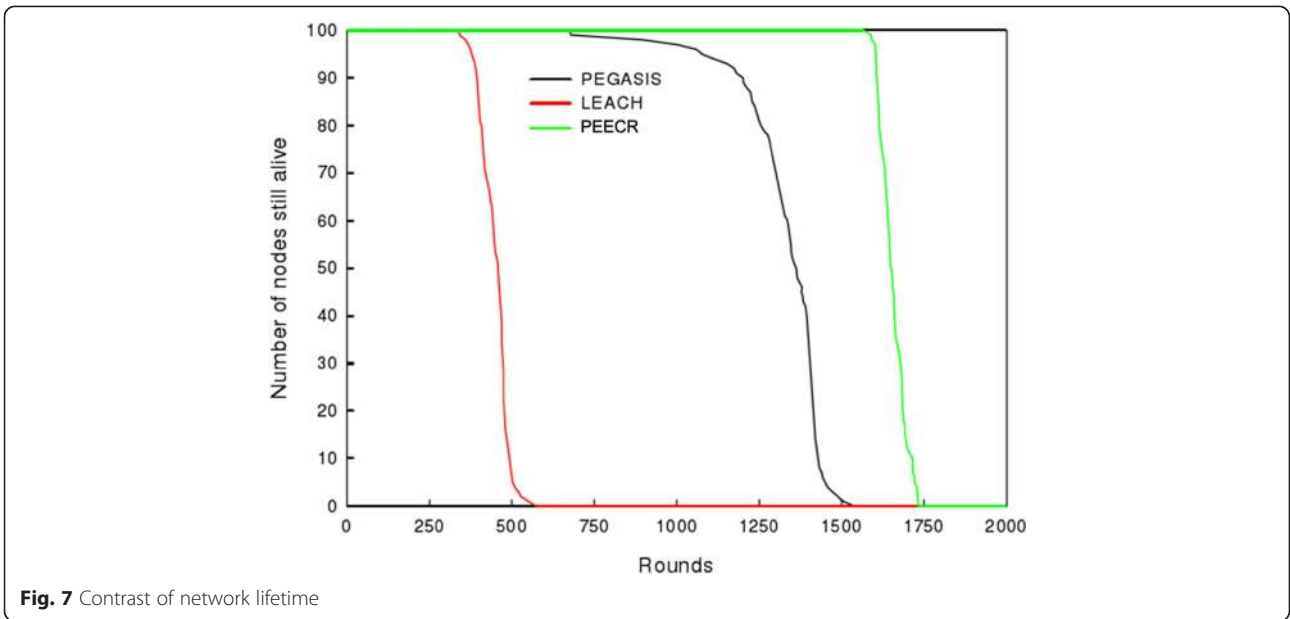**Fig. 6** Analysis of relationship between the number of clusters and the survival time

Zhang *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:162

Page 11 of 13

Fig. 7 Contrast of network lifetime

appears at the 554th round, which is 0.34552 J, and the lowest at the 533rd round, which is 0.0407 J. That is to say that the network energy consumption of each data communication cycle is fluctuating in a larger energy region between 0.0407 and 0.34552 J; it indicates that the LEACH protocol makes the energy consumption of network so non-uniform, so after running up for short 683 rounds, nodes start to die one after another. After the 967th round, there is also much residual energy in the network, but the network was almost unable to communicate effectively at this time. Relative to the LEACH protocol, the network using PEGASIS has the better

balance of energy consumption: the average energy consumption in the forgoing 1214 rounds is 0.13156 J, and it is fluctuating in an energy region between 0.096673 and 0.214323 J.

As the balance of energy consumption is better than that with the LEACH protocol, so the network using the PEGASIS protocol can serve for a longer time. The PECE routing protocol proposed in this paper shows the best performance in the test: in the forgoing 1649 rounds of running network, the shocks of energy consumption of rounds are all small except some individuals, and the values are all close basically to the average
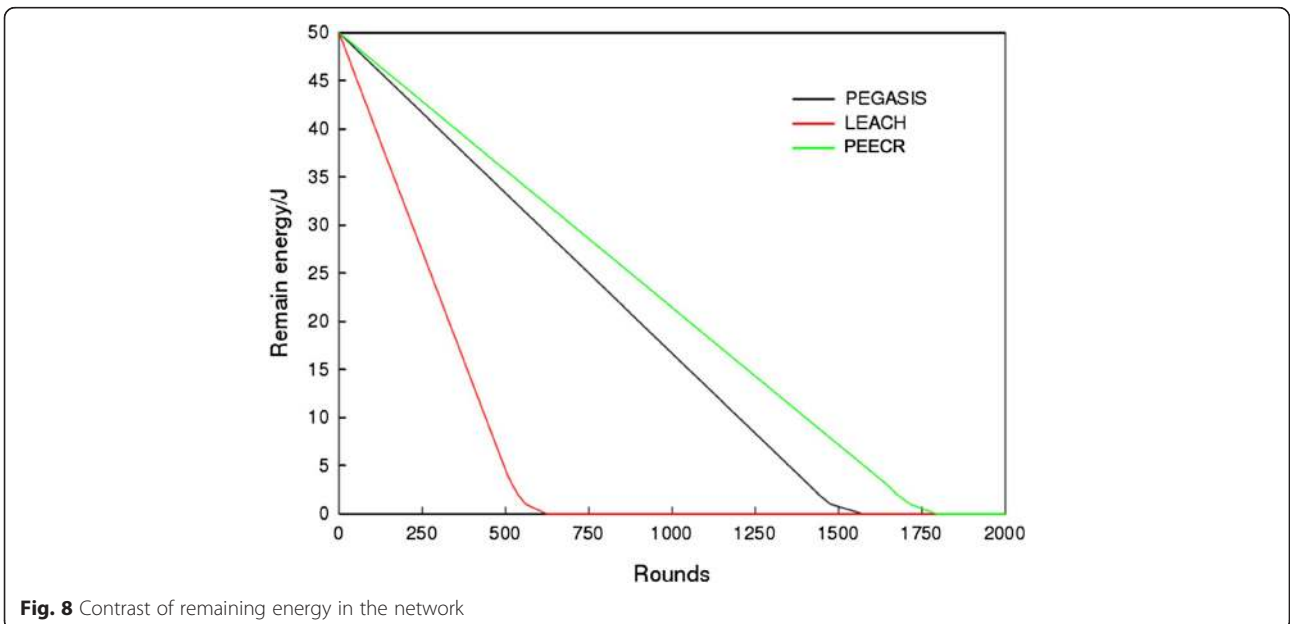

Fig. 8 Contrast of remaining energy in the network

Zhang *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:162

Page 12 of 13

energy consumption of 0.08809 J. The PECE routing protocol which is based on the node degrees, the relative distance between nodes, and the residual energy of nodes has shown well performance in energy-efficient clustering algorithms. Also, the strategy for predictive energy-efficient data transmission based on the BCO model saves a considerable amount of energy for data communication and performs excellently.

6 Conclusions

We have presented a new clustering routing method based on PECE for WSN in this paper. It consists of two stages: cluster formation and stable data transfer. In the first stage, this method chooses optimally the nodes which have more neighbor nodes and shorter relative distance away from neighbor nodes to be the cluster head. Meanwhile taking into account the residual energy of nodes, the-low energy nodes are not likely to become the cluster head nodes. The formed clusters by such method, not only the number of cluster heads and the cluster size, have been controlled, but also reducing the cost of communications between member nodes and cluster head nodes, improving the quality of the cluster, thereby increasing the performance of the overall network. In the second stage, by using bee colony optimization, we design a new strategy for predictive energy-efficient data transmission. On the basis of considering the predictive values of energy consumption, the hops, and the propagation delay on this route, this strategy gives a precise definition of the route yield by using two types of bee agent to predict the route yield of each routing paths from the source node to the sink node. The experimental results show that the PECE algorithm not only significantly reduces the energy consumption of the network but also improves the balance of network energy consumption and extends efficiently the lifetime of the network, and compared with traditional clustering routing algorithm, the method has obvious advantages.

Author details
[1]Key Laboratory of Computer Vision and System (Tianjin University of Technology), Ministry of Education, Tianjin 300384, China. [2]Tianjin Key Lab of Intelligent Computing and Novel Software Technology, Tianjin University of Technology, Tianjin 300384, China. [3]Department of Computer Science and Technology, Tangshan College, Tangshan 063200, China.

References
1. DG Zhang et al., Design and implementation of embedded un-interruptible power supply system (EUPSS) for web-based mobile application. Enterprise Inform Syst **6**(4), 473–489 (2012)
2. DG Zhang et al., A novel multicast routing method with minimum transmission for WSN of cloud computing service. Soft Computing (2014). doi:10.1007/s00500-014-1366-x
3. GQ Zheng et al., The research process of MAC protocol of WSN. Acta Automatic Sinica **34**(3), 305–316 (2008)
4. S Yi et al., PEACH: power-efficient and adaptive clustering hierarchy protocol for wireless sensor networks. Comput. Commun. **30**, 2842–2852 (2007)
5. J Kennedy et al., Particle swarm optimization. Proceeding of IEEE International Conference Neural Networks **4**(1), 1942–1947 (1995)
6. GG Llinas et al., Network and QoS-based selection of complementary services. IEEE Trans. Serv. Comput. **8**(1), 79–91 (2015)
7. DG Zhang, A new approach and system for attentive mobile learning based on seamless migration. Appl. Intell. **36**(1), 75–89 (2012)
8. S Stanislava et al., Cluster head election techniques for coverage preservation in wireless sensor networks. Ad Hoc Netw. **5**(7), 955–972 (2009)
9. I Fahmy, et al., Evaluating Energy Consumption Efficiency of the Zone-Based Routing Protocol. Proceedings of the 46th conference for statistics, computer sciences and operations research, December 2012, Giza, Egypt (Springer, Heidelberg, Germany).
10. DG Zhang et al., Novel Quick Start (QS) Method for Optimization of TCP. Wirel. Netw **5**, 110–119 (2015). doi:10.1007/s11276-015-0968-2
11. CY Li et al., The study and improvement of LEACH in WSN. Chin J Sensors Actuators **23**, 1163–1167 (2010)
12. S Lindsey et al., PEGASIS: power efficient gathering in sensor information systems. IEEE Aero Conf **3**, 1152–1130 (2002)
13. F Huang et al., Energy consumption balanced WSN routing protocol based on GASA. Chin J Sensors Actuators **4**, 586–592 (2009)
14. YA Ridhawi et al., Decentralized plan-free semantic-based service composition in mobile networks. IEEE Trans. Serv. Comput. **8**(1), 17–31 (2015)
15. D Karaboga et al., A novel clustering approach: artificial bee colony (ABC) algorithm. Appl. Soft Comput. **2011**(11), 652–657 (2011)
16. DG Zhang et al., A new method of constructing topology based on local-world weighted networks for WSN. Comp Math App **64**(5), 1044–1055 (2012)
17. DG Zhang et al., A kind of novel method of service-aware computing for uncertain mobile applications. Math. Comput. Model. **57**(3–4), 344–356 (2013)
18. IK Samaras et al., A modified DPWS protocol stack for 6LoWPAN-based wireless sensor networks. IEEE Trans Ind Inf **9**(1), 209–217 (2013)
19. DG Zhang et al., An energy-balanced routing method based on forward-aware factor for wireless sensor network. IEEE Trans Ind Inf **10**(1), 766–773 (2014)
20. DG Zhang et al., A novel approach to mapped correlation of ID for RFID anti-collision. IEEE Trans Serv Comp **7**(4), 741–748 (2014)
21. M Li et al., A survey on topology control in wireless sensor networks: taxonomy, comparative study, and open issues. Proc. IEEE **101**(12), 2538–2557 (2013)
22. DG Zhang et al., A novel image de-noising method based on spherical coordinates system. EURASIP J Ad Sign Proc **1**, 110 (2012). doi:10.1186/1687-6180-2012-110
23. YJ Yao et al., EDAL: An Energy-Efficient, Delay-Aware, and Lifetime-Balancing Data Collection Protocol for Wireless Sensor Networks. MASS **1**(1), 182–190 (2013)
24. Y J Yao, et al., EDAL: An Energy-Efficient, Delay-Aware, and Lifetime-Balancing Data Collection Protocol for Heterogeneous Wireless Sensor Networks. (2014) doi:10.1109/IEEE/ACM Transactions on Networking.2014.2306592
25. YY Zeng et al., Real-time data report and task execution in wireless sensor and actuator networks using self-aware mobile actuators. Comput. Commun. **36**(9), 988–997 (2013)
26. DG Zhang et al., A new anti-collision algorithm for RFID tag. Int. J. Commun. Syst. **27**(11), 3312–3322 (2014)
27. DJ He et al., ReTrust: attack-resistant and lightweight trust management for medical sensor networks. IEEE Trans. Inf. Technol. Biomed. **16**(4), 623–632 (2012)
28. DJ He et al., A distributed trust evaluation model and its application scenarios for medical sensor networks. IEEE Trans. Inf. Technol. Biomed. **16**(6), 1164–1175 (2012)

Zhang *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:162

Page 13 of 13

29. YY Zeng et al., Directional routing and scheduling for green vehicular delay tolerant networks. Wirel. Netw **19**(2), 161–173 (2013)
30. YN Song et al., A biology-based algorithm to minimal exposure problem of wireless sensor networks. IEEE Trans. Netw. Serv. Manag. **11**(3), 417–430 (2014)
31. YS Yen et al., Flooding-limited and multi-constrained QoS multicast routing based on the genetic algorithm for MANETs. Math. Comput. Model. **53**(11–12), 2238–2250 (2011)
32. Z Sheng et al., A survey on the ietf protocol suite for the internet of things: standards, challenges, and opportunities. IEEE Wirel. Commun. **20**(6), 91–98 (2012)
33. G Acampora et al., A survey on ambient intelligence in healthcare. Proc. IEEE **101**(12), 2470–2494 (2013)
34. AV Vasilakos et al., *Delay Tolerant Networks: Protocols and Applications. CRC Press* (Taylor & Francis Group, London, England, 2012)
35. Y Xiao et al., Tight performance bounds of multihop fair access for MAC protocols in wireless sensor networks and underwater sensor networks. IEEE Trans. Mob. Comput. **11**(10), 1538–1554 (2012)
36. X Liu et al., Compressed data aggregation for energy efficient wireless sensor networks. SECON **1**(2), 46–54 (2011)
37. HJ Cheng et al., Nodes organization for channel assignment with topology preservation in multi-radio wireless mesh networks. Ad Hoc Netw. **10**(5), 760–773 (2012)
38. S Sengupta et al., An evolutionary multiobjective sleep-scheduling scheme for differentiated coverage in wireless sensor networks. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **42**(6), 1093–1102 (2012)
39. GY Wei et al., Prediction-based data aggregation in wireless sensor networks: combining grey model and Kalman filter. Comput. Commun. **34**(6), 793–802 (2011)
40. M Chen et al., Body area networks: a survey. Mobile Net App **16**(2), 171–193 (2011)
41. XF Wang et al., A survey of green mobile networks: opportunities and challenges. Mobile Net App **17**(1), 4–20 (2012)
42. P Li et al., CodePipe: an opportunistic feeding and routing protocol for reliable multicast with pipelined network coding. INFOCOM **1**(1), 100–108 (2012)
43. L Liu et al., Physarum optimization: a biology-inspired algorithm for the Steiner tree problem in networks. IEEE Trans. Comput. **64**(3), 819–832 (2015)
44. Y Liu et al., Multi-layer clustering routing algorithm for wireless vehicular sensor networks. IET Commun. **4**(7), 810–816 (2012)
45. C Busch et al., Approximating congestion + dilation in networks via "quality of routing" games. IEEE Trans Comp **61**(9), 1270–1283 (2012)