http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 8 (1): 28-38, 2009
ISSN 1812-5638

An Energy Efficient and Balance Hierarchical Unequal Clustering Algorithm for Large Scale Sensor Networks

Y. Wang, T.L.X. Yang and D. Zhang
School of Electronics and Information Engineering, Xi'an Jiaotong University,
Xi'an, Shaanxi, People's Republic of China

Abstract: Organizing Wireless Sensor Networks (WSN) into clusters enables the efficient utilization of the limited energy resources of the deployed sensor nodes. However, the problem of unbalanced energy consumption exists and it is tightly bound to the role and to the location of a particular node in the network. The so-called hot spot occurs when cluster heads closer to the sink node are burdened with heavier relay traffic and tend to die much faster. To mitigate or avoid the problem, the Partition Energy Balanced and Efficient Clustering Scheme (PEBECS) has been proposed, which divides the entire WSN into several equal partitions reasonably and groups the nodes into clusters of unequal sizes. Cluster heads in these partitions closer to the sink node have smaller cluster sizes than those farther, thus they can preserve some energy for the inter-cluster communications. Further, the cluster heads are elected by using a node-weight heuristic algorithm, where the node's residual energy, the node's degree difference and the relative location in WSN are considered, such that more balanced load is achieved. Simulation results show that PEBECS outperforms significantly in optimizing the cluster heads' energy consumption, balancing the nodes' energy consumption, prolonging the network lifetime and improving the network scalability.

Key words: Wireless sensor networks, clustering, energy efficiency, network lifetime, hot spot

INTRODUCTION

Lifetime of sensor nodes determines lifetime of a Wireless Sensor Networks (WSN), which is essential for various applications. One of the most restrictive factors on the lifetime of WSN is the limited energy resources of the deployed sensor nodes. In order to achieve high energy efficiency and assure long network lifetime, sensor nodes can be organized hierarchically by grouping them into clusters, where data is collected and processed locally at the Cluster Heads (CHs) before being sent to a Sink Node (SN). The clustering algorithm should also be effective in increasing network scalability (Akyildiz *et al*., 2002; Al-Karaki and Kamal, 2004).

During the last few years, many clustering algorithms have been proposed as an efficient way to organize communication and data processing in WSN (Bonivento *et al*., 2007; Chong *et al*., 2007; Lindsey *et al*., 2002; Ming *et al*., 2007; Muruganathan *et al*., 2005; Zhi *et al*., 2007). LEACH (Heinzelman *et al*., 2002), which is the first clustering protocol, proposes a two-phase mechanism based on single-hop communication. The plain node transmits the data to the corresponding cluster head and the cluster head transmits the aggregated data to the sink node. HEED (Younis and Fahmy, 2004) selects cluster heads through O(1) time iteration according to some metric and adopts the multi-hop communication to further reduce the energy consumption. PEGASIS (Lindsey and Raghavendra, 2002) improves the performance of LEACH and prolongs the network lifetime greatly with a chain topology. Although some energy is saved, the resulting delay is significant. In EARACM (Sangho *et al*., 2005), some overhearing nodes are selected as relay nodes and adaptive clustering mechanism is used for routing.

However, two challenges remain in the design of clustering schemes, namely:

- How many clusters are required to be created?
- How should the clusters be formed?

To answer the two questions, one important problem of clustering algorithms clearly, which is called the hot spot (Ming *et al*., 2007; Muruganathan *et al*., 2005; Sevgi and Kocyigit, 2008), should be understood. The result is that, CHs closer to the SN are more active, serving as relay stations for messages coming from upper partitions of the network, which creates unbalanced energy consumption among the cluster head nodes.

Corresponding Author: Yi Wang, School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, People's Republic of China

As a novel solution to the first problem, an approach, where the whole WSN is divided into several equal partitions, the different number of nodes in each partition is elected as CHs and the nodes are organized into clusters of different sizes, should be investigated and analyzed. In general, the energy consumed on intra-cluster communication changes proportionally with the number of nodes within a cluster, while the energy spent on inter-cluster communication is a function of the expected load from the clusters further away. Therefore, by changing the number of nodes in every cluster with respect to the expected relay load, more uniform energy consumption among the CHs can be maintained, so that the total energy dissipated for every CH is similar. To achieve this goal, after employed present algorithm, the result is that clusters in a partition closer to the SN are expected to have smaller cluster sizes, while clusters in a partition farther away have larger sizes.

The second question includes two aspects: how to select the CHs and how to associate a non-CH node to a particular cluster. The approach based on node weight is investigated to electing the CHs, which assigns node weights according to the suitability of nodes acting as CHs and election of the CH is done on the basis of the largest weight among its neighbours. This new mechanism is based on the use of a combined weight metric, which takes into account several system parameters. In order to equalize the energy consumption among nodes, it is imperative that CHs are changed several times during the lifetime of the network.

In clustering algorithms, the energy dissipation for a given sensor node is governed by two main factors, the role that a sensor node takes during each round and the relative location of a sensor node. Once a node elected as CH, it has a competition range to attract the non-CH node to join. Different competition ranges are used to produce clusters of unequal sizes in each partition. In order to mitigate the hot spot problem, because of the influence of node's relative location, even in a same partition, the cluster sizes are different.

It is so interested in exploring a deterministic approach, where the number of elected CHs varying according to the partition near or far away from the SN, managing at the same time the size of their clusters and the expected load from other clusters further away. The problem of unbalanced energy consumption, particularly among the CHs, should be solved. As a new approach to clustering the network, a proposed scheme, which is referred to as Partition Energy Balanced and Efficient Clustering Scheme (PEBECS), where the clusters' sizes (and therefore the number of nodes in every cluster), are determined in a way such that more balanced energy consumption among the CHs is achieved. Figure 1 shows the data flow in a clustered network by employing the PEBECS algorithm.

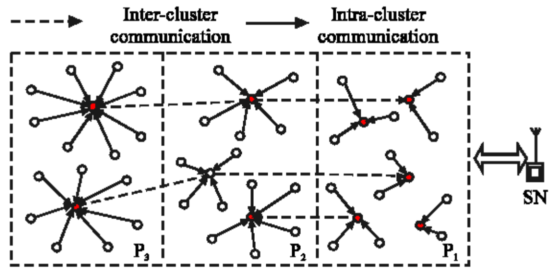

Fig. 1: The data flow in PEBECS algorithm

PRELIMINARIES

The important operation in sensor node clustering is to elect a set of CHs among the nodes in the networks and cluster the rest of the nodes with these heads. CHs are responsible for coordination among the nodes within their clusters (intra-cluster coordination) and communication with each other and/or with external SN on behalf of their clusters (inter-cluster communication). A few assumptions about the network model have been made and the communication model and data aggregation model before the problem statements have been introduced.

Network model: PEBECS divides the operation process into the network clustering phase T_c, be the time spent to cluster the network and the data gathering phase T_d be the time interval between the end of a T_c interval and the start of the subsequent T_c interval. Generally, $T_d >> T_c$. PEBECS ensures that to reduce the overhead.

A set of sensors are assumed to be distributed densely on a flat two-dimension field. The following assumptions of the properties of WSN can be presented:

- The sensor nodes are stationary and homogeneous
- Only one sink node exists and is located outside the deployment area
- Each node has a fixed number of transmission power levels and uses different power levels to communicate within or across the clusters
- A CH is able to aggregate data from the member nodes in same cluster; at the meanwhile, it relays the data from other CHs directly instead of aggregating the data
- Node communications with its CH directly; CH uses multi-hop communication with the SN
- Sensor node is location-unaware; however, each node holds a partition flag bit (PID)

In this study, two methods are proposed for sensor node to obtain the partition flag bit PID. If the WSN is deployed manually, PID is arranged in default; otherwise, if the WSN is deployed randomly, PID can be achieved by adopting the RSSI-based method (Bahl and Padmanabhan, 2000; Bonivento *et al*., 2007). In the networks configuration, the SN broadcasts a series of beacon packets and sensor node receives these packets, records and measures the signal strength to estimate the relative distance between it and the SN.

Wireless communication energy consumption model: The following summarizes the energy consumption model for each sensor component. The key energy parameters for communication in this model are the energy consumed per transmitted bit by the transmitter E_t, energy dissipated per bit in the process of transmission E_α and energy consumed per bit by the receiver electronics E_r. Assume that the total energy consumption is E_c per processed bit by the computation devices and a packet is l b. Depending on the transmission distance both the free space e_n and the multi-path fading e_f channel models are used. Thus, if a node transmits l bits message through a distance d, the energy $E_T(l,d)$ it expends:

$$E_T(l,d) = E_t(l) + E_c(l) + E_\alpha(l,d) \begin{cases} lE + le_n d^2 & (d \le d_0) \\ lE + le_f d^4 & (d > d_0) \end{cases} \quad (1)$$

and to receive the message, the energy $E_R(l)$ expended:

$$E_R(l) = E_r(l) + E_c(l) = lE \quad (2)$$

Data aggregation model: The CHs were assumed by Chong *et al*. (2007), Lindsey *et al*. (2002), Seung *et al*. (2004) and Younis *et al*. (2006) to be able to perfectly aggregate multiple incoming packets into one outgoing packet. However, it is applicable to those cases when the sensed data are correlated with each other. In this model, some certain aggregation algorithms are not employed, however, a particular coefficient, called aggregation coefficient α ($\alpha \in [0, 1]$), is used to represent the function of those algorithms. Let $\alpha = 1$ indicate the case of perfect aggregation, while $\alpha = 0$ means that CHs do not conduct any aggregation.

Problem statement: Periodical data gathering applications in large scale sensor networks appeal the design of scalable, energy efficient clustering algorithms. Network lifetime can be defined as the time elapsed until the first node in the network depletes its energy. Once a sensor node runs out its energy, the network could be considered that it is dead because some area cannot be monitored any more. The CHs using multi-hop communication with the SN leads to hop spot problem in the network, where, CHs in the hot spot use their energy at a much higher rate and die much faster than the others. Mitigating or avoiding the problem becomes necessary in order to prevent the problem of premature battery drainage for these CHs near the SN. Thus, PEBECS aims at (1) balancing nodes' energy consumption and prolonging the sensor networks' lifetime, (2) improving the scalability of the WSN and (3) decreasing the clustering cost and balancing the traffic load.

PEBECS ALGORITHM

In this study, assume that N nodes are dispersed in a field and the entire sensing area of the WSN is divided into K local partitions. K is a system parameter. These partitions are denoted as $P_1, P_2, ..., P_K$. P_1 is the nearest partition from the SN and P_K is the farthest. In each partition, n = N/K nodes are deployed and each partition area is $S_p = S/K$. Each partition i has a certain number of clusters O_i. Further, each node stores a partition flag bit PID, which is used to indicate which partition it belongs to.

There are two types of CHs in WSN. A CH in P_K is different from a CH in other partitions. The key point is that a CH in P_K is not responsible for relaying the data, while the CHs in other partitions take on the heavy burden of relaying these up and down messages. A CH in $P_1, P_2, ..., P_{K-1}$ is regarded as full-duty CH, while a CH in P_K as part-duty CH. PEBECS forms clusters by using a fully distributed algorithm which is conducted in the following two phases: cluster head election and cluster formation.

Cluster head election: In this phase, several suitable sensor nodes in each partition are elected as CHs. PEBECS, which is based on node weight, can effectively elect the proper nodes as CHs by combining each of the necessary system parameters with certain weighting factors chosen according to the application requirements. This means that a node decides to become a CH or stay as a CM depending on its combined weight metric, which takes into account the parameters like the energy threshold ratio Δ_E, the degree difference ξ, the No. of nodes in the neighbor partitions γ and the node No. difference Δ_n.

```
Begin Procedure
A sensor node m in Partition p:
1. Form HI_Adv_Msg(ID,E_residual(m),PID)
2. Broadcast Msg(HI_Adv_Msg)
3. Receive HI_Adv_Msg_Neighbour_Info_Table
4. Begin
5. For each node i1 in Neighbour_Info_Table
6. E_ave= E_ave+Neighbour_Info_Table[i1].E_residual(i1)
7. End For
8. E_ave = E_ave/Neighbour_Info_Table.length
9. If (E_residual(m)<E_ave) Exit
10.Else ΔE= (E_residual(m)-E_ave)/(E_residual(m)-E_ave)
11. Begin
12. For each node i2 in Neighbour_Info_Table
13. If (Neighbour_Info_Table[i2].
                              PID == p-1)
14. n_p-1++;
15. Else If (Neighbour_Info_Table[i2].
                              PID == p+1)
16. n_p+1++;
17. Else Continue
18. End For
19. ϵ=|n/O_p- Neighbour_Info_Table.length|
20. γ=(n_p+1+n_p-1)
21. Δn=|n_p+1-n_p-1|
22. W_m=w1xΔE + w2xϵ + w3xγ - w4xΔn
23. Broadcast Node_Weight(W_m,ID,PID)
24. Receive Node_Weight(W_Node_Weight_Table)
25. Begin
26. For each node weight i3 in Node_Weight_Table
27. If (W_m < Node_Weight_Table[i3])
28. Is_Cluster_Head = False
29. Exit
30. End If
31. End For
32. If(Is_Cluster_Head)
33. Mark node m as Cluster_Head
34. Broadcast Msg(CH_Sus_Msg)
35. End If
36. Invoke Cluster Radius Competition
End Procedure
```

The pseudo code of the cluster head election is shown above. The first component at line 10, Δ_E, is mainly related to sustainable energy consumption. The more residual energy a node remains, the higher probability to be a CH. Since the No. of nodes that a CH in partition P_i can handle ideally is n/O_i, the second component at line 19, ξ, is used to ensure that CHs are not over-loaded and the efficiency of the system is maintained at the expected level. Unlike LEACH in which the CHs' location is ignored, PEBECS introduces the nodes' relative location metric, which includes the third and fourth components, γ and Δ_n. The motivation of this metric is that it helps in efficient energy saving because it is always desirable for a CH to be at right location. The third component $w_3(n_{p+1}+n_{p-1})$ helps PEBECS prefer to choose a node that communicates with as many nodes as possible in neighbor partitions within its maximum transmission range. When the nodes' residual energy and the number of nodes in neighbor partitions are even or approximate, the fourth component $w_4\Delta_n$ prevents a node which is very near or far from the other neighbor partitions from being elected as a CH. However, it should be taken into consideration that the residual energy will be more for nodes acting as CHs. This influence is embodied by the weight factors for the corresponding parameters. Further, the requirements of reality applications are taken into consideration when choosing eligible w_1, w_2, w_3 and w_4 The flexibility of changing the weight factors helps us apply present algorithm to various network applications.

The output of cluster head election procedure is a set of nodes called the CH set in each partition. The cluster head election is invoked at the time of system activation and also when a CH node in the current CH set is unable to act as CH or its energy drainage rate is excessively high. Every invocation of the election algorithm does not necessarily mean that all the CHs in the earlier set are replaced with the new ones. If a node detaches itself from its current CH and attaches to another CH then the involved CHs update their member list instead of invoking the election algorithm.

Cluster formation: Here, in order to mitigate the hot spot problem and balance CH's energy consumption, a cluster radius competition algorithm is introduced, which aims to assign the No. of cluster member nodes (CMs) within a cluster closer to SN to be smaller, even when these clusters are in same partition. The cluster competition radius R_C is defined as:

$$R_C=\left[1-\delta\left(1-\frac{d(ch_m,ch_n)}{\Phi}K\right)\right]R_T \quad (3)$$

where, δ is a system parameter whose value is defined by actual application, in the simulations $\delta = 2/7$, K is the partition number, Φ is the diameter of the entire WSN, R_T is the maximum of sensor node's transmission range and $d(ch_m,ch_n)$ is the distance between CH ch_m in partition p and CH ch_n in neighbor partition p-1; especially, if p = 1, ch_n = SN.

```
Begin Procedure
A   CH node ch in partition p:
Form Competition_Msg(ID,PID,R_C)
Broadcast Msg(Competition_Msg)
A   sensor node m:
Receive Competition_Msg C1, C2,…,Cl
Begin
To Competition_Msg Cn (n∈[1,l])
If(Cn.PID = m.PID)
Cn→m.Competiton_info_Table
Else If
Abandon Cn
For each Msg i in m.Competiton_info_Table
Ck.RC = = min(i.RC)
```

```
12. End For
13. CM m∈CH k
14. CM m sends Adhere_Msg(ID,CHID(l))
15. CH k accepts Adhere_Msg(ID,CHID(l))
16. m.ID→k.Member_info_Table
17. End
End Procedure
```

The pseudo code of the cluster formation is shown above. After elected, CHs broadcast Competiton_Msg within maximum radio range R_C to advertise their wills (line 1-2). Each CM node receives some pieces of Competiton_Msg C_1, C_2,...,C_l and checks whether the PID of Competiton_Msg is equal to its PID (line 3-10). Assume that there is a CM node m, m always finds the minimum of R_C and will join that cluster (line 11-16).

THEORETIC ANALYSIS

Here, the performance of PEBECS will be discussed in details and the optimal number of clusters in each partition and several constraints of PEBECS, such as the network delay, the admission degree restriction in a MAC cycle, are discussed.

Theorem 1: The message complexity of PEBECS is *O*(N), where, N is the number of nodes in the entire WSN.

Proof: Observing PEBECS, in the initialization phase, every node broadcasts HI_Adv_Msg which is received by all other nodes lying within its transmission range. Some nodes quit the algorithm execution because the residual energy is lower than the mean value of the residual energy E_{ave} which can be estimated according to the received messages. It is assumed that the portion of these node is $(1-\beta)N$ $(\beta \in [0, 1))$ and the number of the other nodes is βN. Then these nodes broadcast the cluster head election weight information message and receive this kind of message from the neighbor nodes. A node, if its weight W is more than the others, is elected as a CH and it broadcast the message CH_Sus_Msg, otherwise, it quits the election. In all K partitions $\sum_{i=1}^{K} O_i$, nodes are elected as cluster head, thus $\sum_{i=1}^{K} O_i$, Competition_Msg messages are broadcasts to cluster the WSN, at the meanwhile, $(N-\sum_{i=1}^{K} O_i)$ Adhere_Msg messages are sent out. Through analysis, the message complexity of PEBECS is given by:

$$N+\beta N+\beta N+\sum_{i=1}^{K} O_i + N - \sum_{i=1}^{K} O_i = 2(1+\beta)N \tag{4}$$

Clearly, the asymptotic order is O(N).

Theorem 2: The time overhead of PEBECS algorithm is mainly related to the topology of the WSN.

Proof: An interesting result has been proposed by Younis *et al*. (2006), is that the time complexity of the cluster algorithm in wireless networks is bounded by a network parameter that depends on the network topology rather than on the size of the network, i.e., the invariant number of the nodes. As mentioned earlier, in order to decide whether it is going to be a CH or a CM, each node waits for the decision of all the neighbor nodes with bigger weight in same partition. This waiting time of each node can be defined as a function of the distance of the two farthest nodes in same partition. Clearly, this distance D_w depends on the current topology of the network rather than on the No. n of the nodes. By the total ordering induced on the nodes by their weights and through a simple inductive argument, it can easily get the result. Each node of the network sends out exactly one message within $D_w + 1$ steps.

Optimal No. of Clusters in each partition: As mentioned earlier, a CH node in partition P_K is part-duty CH while in other clusters are full-duty CH. Thus, a CH in partition P_K consumes energy by receiving from the member nodes, aggregating and transmitting the message to next hop CH in partition P_{K-1}. Therefore, the energy dissipated by a CH in partition P_K in T_d phase is:

$$\begin{aligned} E_{CH(K)} &= l\left[E_R\left(\frac{n}{O_K}-1\right)+E_{DA}\frac{n}{O_K}+E_T\frac{\alpha n}{O_K}\right] \\ &= l\left[E\left(\frac{n}{O_K}-1\right)+E_{DA}\frac{n}{O_K}+\left(E+e_f D^4\right)\frac{\alpha n}{O_K}\right] \end{aligned} \tag{5}$$

where, D is the distance between this CH and next hop CH and α is the data aggregation coefficient, E_{DA} is the energy consumed to aggregate a message by CH.

Each CM node only needs to communicate with and transmit message to its CH. The energy consumption of a non-CH node in partition P_K is:

$$E_{non\text{-}CH(K)} = lE_T = l\left(E+e_n d_K^2\right) \tag{6}$$

The area occupied by each cluster is approximately $S_p/O_K = S/KO_K$. The expected squared distance from the nodes to CH can be given by:

$$E\left(d^2\right)=\iint\left(x^2+y^2\right)\rho(x,y)\,dxdy=\iint r^2\rho(r,\theta)\,drd\theta \tag{7}$$

The area is regarded as a circle area with radius $R = (S/\pi KO_K)^{1/2}$ and $\rho(r, \theta)$ is constant for r and θ. The equation above can be simplified to:

$$E\left(d_K^2\right)=\int_0^{2\pi}\int_0^{R}\rho r^3 dr d\theta=\frac{\rho S^2}{2\pi K^2 O_K^2} \quad (8)$$

If assume that the node deployment in area S is uniform, then ρ = N/S = Kn/S. Therefore, Eq. 8 can be expressed as:

$$E\left(d_K^2\right)=\frac{\rho S^2}{2\pi K^2 O_K^2}=\frac{KnS^2}{2\pi SK^2 O_K^2}=\frac{nS}{2\pi K O_K^2} \quad (9)$$

In this case, the energy dissipated by all nodes in a cluster during a complete process is:

$$E_{cluster}=E_{CH(K)}+\left(\frac{n}{O_K}-1\right)E_{non\text{-}CH(K)} \approx E_{CH(K)}+\frac{n}{O_K}E_{non\text{-}CH(K)} \quad (10)$$

and the total energy consumption is:

$$E_K=O_K E_{cluster} \approx nl\left[(2+\alpha)E+E_{DA}+\alpha e_f D_K^4+\frac{e_n nS}{2\pi K O_K^2}\right] \quad (11)$$

where, D_K is concerned with the CH number O_K in this partition and CH nodes' distribution θ, thus D_K is a function of O_K and θ, $D_K = g(O_K, \theta)$, when the case is not perfect data aggregation. To minimize the total energy consumption, the Eq. 11 should be minimized. It is easy to find that by setting the derivative of E_K with respect to O_K to 0. The optimal cluster number is:

$$\widetilde{O_K}=\frac{1}{g(O_K,?)}\left(\frac{e_n nS}{4\alpha\pi e_f K g'(O_K,\theta)}\right)^{1/3} \quad (12)$$

In the case of perfect data aggregation, the total energy consumption of the complete process is:

$$E_K^o=O_K E_{cluster} \approx l\left[2nE+nE_{DA}+O_K e_f D^4+\frac{e_n n^2 S}{2\pi K O_K^2}\right] \quad (13)$$

and the optimal cluster number in partition P_K is:

$$\widetilde{O_K^o}=\left(\frac{n^2 S e_n}{\pi K D_K^4 e_f}\right)^{1/3} \quad (14)$$

If the CH is full-duty CH, the energy dissipated increases because of the energy consumption of relaying data to the CHs in neighbor partitions P_{K-i} ($i\in[1,2,...,K-1]$), the equation of the total partition energy consumption can be computed by:

$$E_{K-i}=2nlE+2\left(\sum_{p=K-i+1}^{K} O_p\right)lE+nlE_{DA}+l\left(\sum_{q=K-i}^{K} O_q\right)e_f D^4+nle_n d_{K-i}^2 \quad (15)$$

Since PEBECS pays attention to balancing the total energy consumption of each partition, the relationship of the total energy consumption of each partition is easily investigated.

$$E_K = E_{K-1} = E_{K-2} = \ldots = E_{K-i} = \ldots = E_1 \quad (16)$$

or, the Eq. 16 can be presented as:

$$E_K = E_{K-i}\,(i\in[1,2,\ldots,K-1]) \quad (17)$$

Using Eq. 11 and 15, after simplified, it is easy to get:

$$O_{K-i}^2\left[2\left(\sum_{p=K-i+1}^{K} O_p\right)E+\left(\sum_{p=K-i+1}^{K} O_p-\widetilde{O_K}\right)e_f D^4-\frac{n^2 S e_n}{2\pi K \widetilde{O_K}^2}\right]+O_{K-i}^3 e_f D^4+\frac{n^2 S e_n}{2\pi K}=0 \quad (18)$$

where, is a cubic equation on the unknown parameter O_{K-i}. Without losing data precision, the mean value of the distance between two neighbor nodes is instead of D in Eq. 18:

$$\overline{D}=\frac{1}{K}\sum_{n=1}^{K} D_K \approx \frac{1}{K}(\Phi) \quad (19)$$

As known, Eq. 18, which is a cubic equation, has two imaginary roots and a real root. Since the cluster parameter in each partition is real number, only the real root must be chosen. For the sake of the conciseness, three labels are defined as follows:

$$A = e_f \overline{D}^4$$

$$B = \left[2\left(\sum_{p=K-i+1}^{K} O_p\right)E+\left(\sum_{p=K-i+1}^{K} O_p-\widetilde{O_K}\right)e_f D^4-\frac{n^2 S e_n}{2\pi K \widetilde{O_K}^2}\right]$$

$$C = \frac{n^2 S e_n}{2\pi K}$$

In Eq. 18, label A, B and C are constant for parameter O_{K-i}. It can thereby be simplified as:

$$AO_{K-i}^3+BO_{K-i}^2+C = 0 \quad (20)$$

The solution of that cubic equation is:

$$O_{K-i}=\frac{1}{6A}\sqrt[3]{4\sqrt{\left(27A^2C+2B^3\right)^2-4B^6}-\left(108A^2C+8B^3\right)}-\frac{B}{3A}-\frac{2B^2}{9A\sqrt[3]{4\sqrt{\left(27A^2C+2B^3\right)^2-4B^6}-\left(108A^2C+8B^3\right)}} \quad (21)$$

Some simulations are conducted to verify the parameter O_i of partition I. The system is simulated of

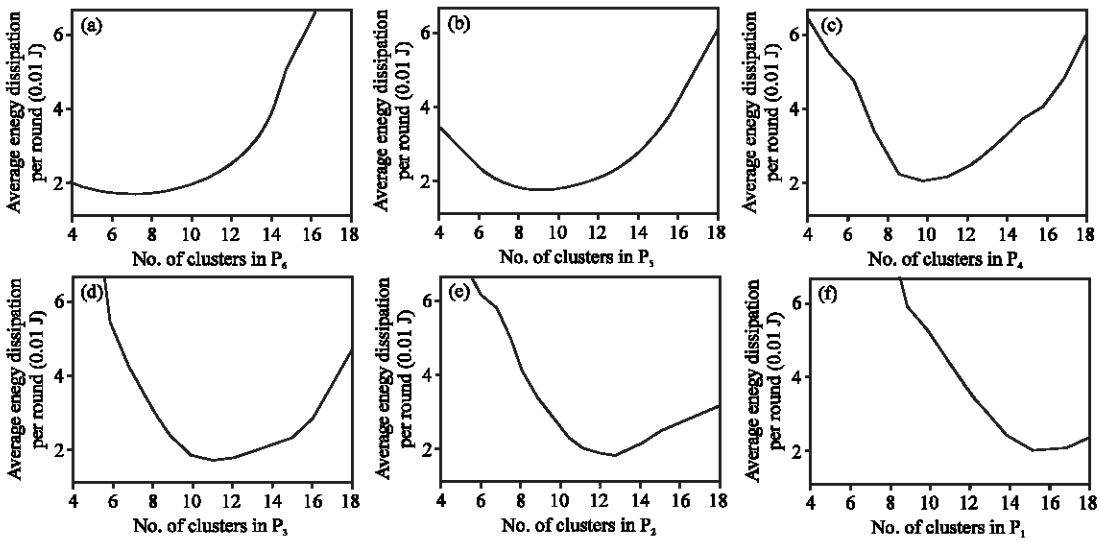

Fig. 2: Optimal number of CHs in each partition, (a) O_6, (b) O_5, (c) O_4,(d) O_3,(e) O_2 and (f) O_1

N = 1200 node, n = 200 node, S = 100×600 m, K = 6 level, e_n = 10 pJ b^{-1} m^{-2}, e_f = 0.013 pJ b^{-1} m^{-4}, E = 50 pJ b^{-1} and 0 m$\le R_T \le$200 m. For simplicity, assume that the probability of signal collision and interference in the wireless channel is negligible. Derived from Eq. 21, the optimal number of CHs in each partition is $O_i \in [4, 19]$ ($i \in [1, 5]$), especially $O_6 \in [6, 18]$.

Figure 2a-f, show the average energy dissipated per round as a function of the number of CHs in each partition. The results are shown for different partitions. According to the analysis above, the optimal number of CHs in partition P_6 should be determined first. Figure 2a shows the average energy consumption per round with respect to the number of CHs in partition P_6. The results are shown for varying O_6. For low number of CHs, the number of nodes in a cluster is relatively high, thus the intra-cluster communication overhead is rise, the CHs in neighbor partitions are hardly find the appropriated relay nodes and some CHs' energy consumption is excessively high. The average energy dissipated per round decreases as the number of CHs increases and reaches a bottom when O_6 is between 5 and 7. Further increase in the number of CHs results in an increase in the energy consumption per round since the number of nodes in a cluster is excessively low, thus the cluster maintenance overhead, clustering speed and energy balance should be concerned. And for the same reason, from Fig. 2, the optimal CH number in each partition are obtained, $O_5 \in [8,10]$, $O_4 \in [9,11]$, $O_3 \in [10,12]$, $O_2 \in [11,13]$, $O_1 \in [15,17]$, respectively. It is interesting to observe all these numbers are in the range of the theoretical results.

Constraints of PEBECS: Some constraints should be taken into account as they help PEBECS better meet the requirements of practical applications in actual world. These constraints include network delay, admissible degree restriction and the impact of partition number.

Generally, network delay t_{ND} consists of two types of delays, system delay t_{SD} and transmission delay t_{ND}. It is difficult to be precise because many possible issues affect the total delay. A typical sensor node includes sensor unit, radio device, memory unit, microprocessor and operating system at least. Sensor unit is responsible for sensing the target and Analog/Digital converting. Microprocessor and memory unit deal with all the data together. Operating system manages data and instructions. The amount of these delays of all three units leads to system delay. On the other hand, the transmission delay consists of the delay of receiving and transmitting data and communication delay. The devices, especially the radio device and the adopted MAC protocol are primary issues.

Because the sensor nodes are homogeneous, the system delay of each node is considered as uniform. Therefore, it can be concluded that the network delay is related with the hop count, through which is the No. of the nodes between the starting node and the SN. Through analysis, the network delay t_{ND} can be defined as a function, whose value depends on the value of independent variables: hop count c_h and the distance of neighbor nodes d_{ij}.

$$t_{ND}(c_h, d_{ij}) = t_{SD}c_h + t_{TD}(c_h, d_{ij}) \quad (22)$$

Therefore, the maximum of t_{ND} can be calculated by:

$$\max t_{ND}(c_h, d_{ij}) = t_{SD}K + t_{TD}(K, \Phi) \quad (23)$$

The second constraint is the admissible degree restriction De, which is defined as a CH node simultaneously manages the maximum of the number of CMs in a same cluster in one MAC cycle. Obviously, MAC protocol is the primary key to De and adopting an appropriate MAC protocol can enlarge De and thus improve the performance. In PEBECS, CHs in the farthest partition have the largest cluster radius; hence they employ the most average No. of CMs in a cluster. That means the admissible degree restriction of CH in the farthest partition always reaches the peak.

The two constraints above limit the range of the value of K. In practical applications, required data freshness guarantees impose timing constraints over the delivery of data from the environment to the end user, thus the network delay should be confined within the scope of reasonable. In a similar way, the MAC protocol not only limits the De of CHs in P_K, but also influences the range of the value of K.

Two simulations are conducted to study the partition number K and the admissible degree restriction De. In the first simulation, N is changed to 2400 and K is varied from 0 to 12. Figure 3 shows the variation of K with respect to the time steps of the first dead CH in WSN occurred. The results are shown for varying K. The time steps of the first dead CH increases as K increases and there is an optimal range for the value of K, about 5-8 in the given scene, then the time steps of the first dead CH decreases as K increases.

In the second simulation, the same network is used and the same MAC protocol TDMA/CSMA in PEBECS as in LEACH and HEED is adapted, however, the probability of signal collision and interference in the wireless channel could not be ignorable. According to the analysis above, the CHs in P_6 are investigated. As the experiment shown in Fig. 4, the lifetime of the first dead CH is longest when the optimal De is about 28-30. This is because, when De is less than 28, the No. of clusters in a partition is larger, thus each CH's energy consumption and the amount of all the cluster maintenance overhead gets quite greater; when De is more than 30, the probability of signal collision and interference in the wireless channel get rather larger, the amount of the intra-cluster communication overhead is relatively larger and the frequency of re-clustering originated gets higher.

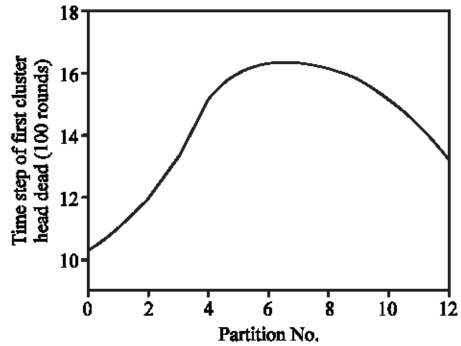

Fig. 3: Rounds of the first CH dead occurred varying K

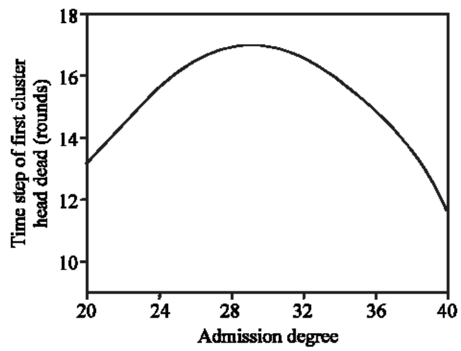

Fig. 4: Rounds of the first CH dead occurred varying De

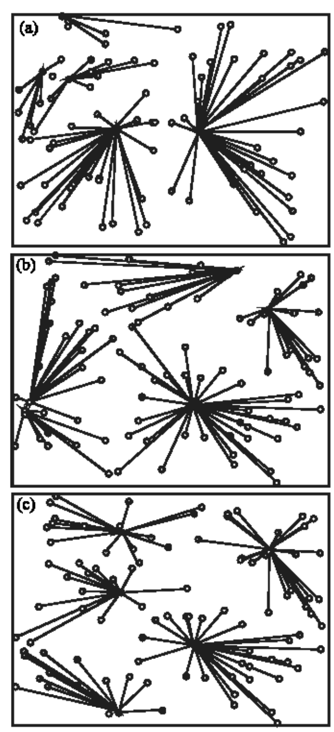

Fig. 5: Example local outputs of the simulation, (a) P_6, (b) P_4 and (c) P_2

In addition, Fig. 5a-c shows the example local outputs of one of the simulations using the PEBECS at some time. Figure 5a-c shows the example local output of partition P_2, P_4 and P_6, respectively. It is interesting to observe that the radius of the clusters are nonequal, specially, the clusters far away from the SN usually consist of many nodes, while the clusters in the near partitions may include few nodes. However, the No. of the clusters in the near partitions is greater.

PERFORMANCE EVALUATION

Here, the performance of PEBECS algorithm is compared with LEACH, HEED, EARACM and PEGASIS based on the simulation results.

LEACH, which is the first clustering protocol, proposes a two-phase mechanism based on single-hop communication. The plain node transmits the data to the corresponding cluster head and the cluster head transmits the aggregated data to the base station (BS). HEED selects cluster heads through O(1) time iteration according to some metric and adopts the multi-hop communication to further reduce the energy consumption. PEGASIS improves the performance of LEACH and prolongs the network lifetime greatly with a chain topology. But the delay is significant although the energy is saved. In EARACM, the algorithm enhances the survivability of networks using adaptive clustering mechanism. The proposed algorithm can reduce the cumulative amount of data packets and ensure efficient use of energy among the nodes in the sensor networks.

In LEACH, since the decision to change the CH is probabilistic, there is a good chance that a node with very low energy gets selected as a CH. When this node dies, the whole cell becomes dysfunctional. Also, the CH is assumed to have a long communication range so that the data can reach the BS from the CH directly. This is not always a realistic assumption since the CHs are regular sensors and the BS is often not directly reachable to all nodes due to signal propagation problems, e.g., due to the presence of obstacles. LEACH also forms one-hop intra and inter cluster topology where each node can transmit directly to the CH and thereafter to the SN. Consequently, it is not applicable to networks deployed in large regions. HEED focuses on the efficient clustering by proper selection of clusterheads based on the physical distance between nodes, but CHs are elected with iteration. Although the communication is localized and the algorithm terminates in O(1) iteration, HEED still produces much more overhead with the upper bound NiterxN. PEGASIS presents the idea that if nodes form a chain from source to SN, only 1 node in any given transmission time-frame will be transmitting to the SN. This mechanism offers promising improvements with relation to network lifetime; however reliability may not be as promising. In PEGASIS, each node communicates with its nearest neighbor. This implementation may be more susceptible to failure due to gaps in the network. Moreover, similar with LEACH, PEGASIS assumes that each node can communicate with the SN directly, which is not applicable. In the setup phase, EARACM uses the localized flooding to find all the routes from source to destination and their energy costs. A destination node initiates a route request and all of the intermediate nodes involve and relay the request in the direction to the source node. The overhead includes packet transmission cost of the advertisement, announcement, joining and scheduling messages from nodes. Furthermore, instead of a single path, a communication would use multiple paths, thus any single path does not get energy depleted, but the overall cost of the network keeps a high level.

In the first simulations, the networks configuration is as follows: N = 600 node, n = 300 node, S = 100×300 m, K = 3, e_n = 10 pJ b^{-1} m^{-2}, e_f = 0.013 pJ b^{-1} m^{-4}, E = 50 pJ b^{-1}, l = 256 b. Note that in LEACH or HEED each node is required to be capable to communication with the SN directly while in EARACM or PEBECS it may not. Therefore, two different communication ranges are introduced: the range in LEACH or HEED is from 0 to 300 m, while in PEBECS or EARACM it is from 0 to 150 m. Lifetime is the criterion for the performance of sensor networks which is determined in terms of round via the ratio of the nodes still alive. Furthermore, the data aggregation coefficient α is not ignorable.

Figure 6a-d show the simulation results for the comparison of the four protocols under the conditions of α varying from 0.2 to 1. As mentioned earlier, when α = 1, it represents the case of perfect aggregation. From Fig. 6d, PEBECS prolongs the lifetime of WSN over 35% against EARACM, 44% against HEED, 51% against LEACH.

From Fig. 6a-c, PEBECS extends the lifetime of WSN by 29, 37, 46% against LEACH, 26, 35, 41% against HEED and 20, 29, 38% against EARACM, in the case of α = 0.2 or 0.5 or 0.8, respectively. This is due to PEBECS always achieves the well distributed CHs with considering not only the residual energy but also the CHs' location; further, to balance the load among the CHs with weighted function is an appropriate method. The results also shows that PEBECS can enlarge the node number of sensor networks effectively and avoid the restriction of distance between starting node to SN that exits in other clustering algorithms.

Another simulation experiment is also performed to evaluate the performance of each protocol (LEACH,

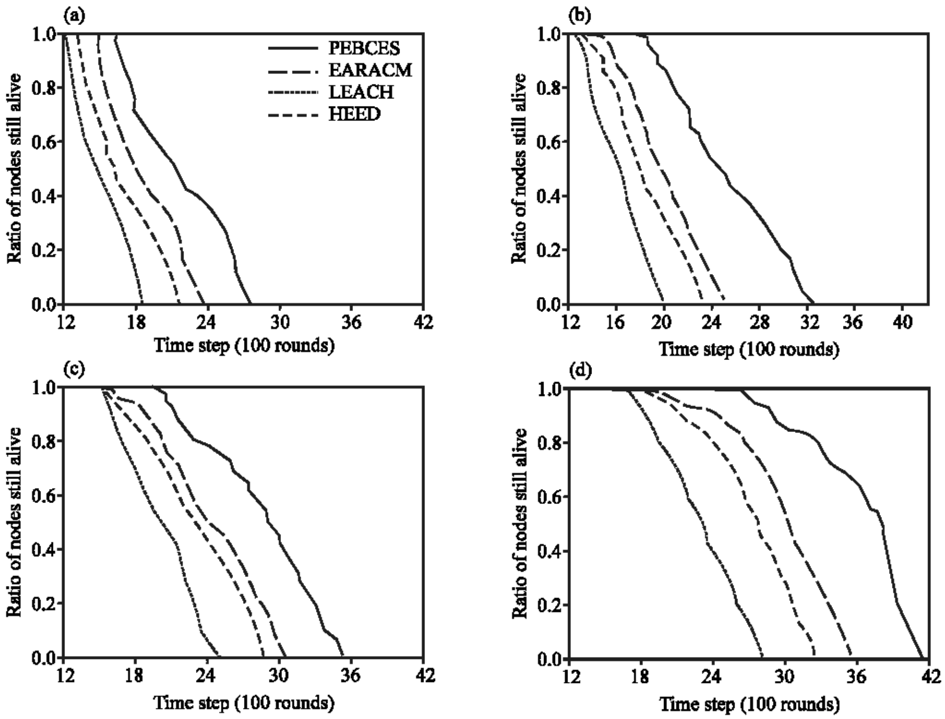

Fig. 6: Performance comparison of the four algorithms, (a) α = 0.2, (b) α = 0.5, (c) α = 0.8 and (d) α = 1

Table 1: Ratio of node alive via different algorithms

	The ratio of node still alive (%)																	
	Round = 1500						Round = 2100						Round = 2800					
Algorithm	P_1	P_2	P_3	P_4	P_5	P_6	P_1	P_2	P_3	P_4	P_5	P_6	P_1	P_2	P_3	P_4	P_5	P_6
LEACH	91.1	88.3	82.5	0.0	0.0	0.0	59.3	48.5	33.8	0.0	0.0	0.0	1.5	1.1	0.8	0.0	0.0	0.0
HEED	95.2	93.6	90.0	0.0	0.0	0.0	69.9	61.2	55.9	0.0	0.0	0.0	10.9	8.4	5.3	0.0	0.0	0.0
EARACM	100.0	100.0	100.0	100.0	98.4	95.2	82.1	78.4	71.2	65.8	61.7	55.6	23.1	18.4	12.3	8.3	1.4	0.0
PEGASIS	100.0	100.0	100.0	99.1	95.6	92.8	79.8	76.4	65.1	61.2	57.8	52.4	27.1	19.9	15.7	9.7	5.5	1.6
PEBECS	100.0	100.0	100.0	100.0	100.0	99.0	89.7	88.2	86.4	85.2	84.3	80.1	57.2	56.1	55.4	53.2	50.7	48.1

HEED, EARACM, PEGASIS and PEBECS) when the amount of the node number in WSN is extended. These parameters can be set $\alpha = 0.8$, N = 600 node, n = 100 node, S = 100×600 m, K = 6 and the communication range in LEACH or HEED is from 0 to 300, while in PEBECS or EARACM it is from 0 to 150. Table 1 shows the ratio of node still alive when LEACH, HEED, EARACM, PGASIS or PEBECS is adopted in WSN. Because of the communication range restriction, there is no node working in partition P_4, P_5 and P_6 when using LEACH or HEED. It is seen from the results that the efficiency of PEBECS performs more distinct when the network scale grows; especially, PEBECS outperforms the others significantly for the nodes in the far partitions.

CONCLUSIONS

In this study, an energy-efficient distributed clustering approach for wireless ad hoc sensor networks has been presented. The hot spot problem arises when employing the multi-hop communication model in a clustered sensor network. In order to mitigate or avoid the problem, the Partition Energy Balanced and Efficient Clustering Scheme, which operates in stationary networks where nodes are location-unaware and have equal significance, has been proposed. The approach is hybrid: the whole WSN is divided into several equal partitions, the different number of nodes in each partition is elected as CHs and the sensor nodes are organized into clusters of different sizes. The cluster heads are randomly selected based on their residual energy, degree difference and relative location in networks and after elected the CHs, the cluster competition radius are adopted to attract the nodes around to join such that energy consumption is minimized and load is balanced.

Simulation results show that PEBECS prolongs network lifetime, improves the network scalability and the clusters it produces exhibit several appealing characteristics. PEBECS parameters, such as the well CH nodes distribution, network scale extension and network operation interval, can be easily tuned to optimize resource usage according to the network deployment and application requirements. The proposed approach can be applied to the design of several types of sensor network protocols that require energy efficiency, scalability, prolonged network lifetime and load balancing. Furthermore, simulation results also show that PEBECS outperforms significantly in optimizing the cluster heads' energy consumption, balancing the nodes' energy consumption and improving the scale of the entire WSN. When different data aggregation strategies are employed by the sensor nodes, the network lifetime is prolonged by 20- 51%.

ACKNOWLEDGMENTS

The authors would like to thank China Next Generation Internet (CNGI 04-1-11D) and National High-Tech and Development Plan of China, under grant 863-2006AA01Z210, for funding this research.

REFERENCES

Akyildiz, I.F., W. Su, Y. Sankarasubramaniam and E. Cayirci, 2002. A survey on sensor networks. IEEE Commun. Maga., 40: 102-114.

Al-Karaki, J.N. and A.E. Kamal, 2004. Routing techniques in wireless sensor networks: A survey. IEEE Pers. Commun., 11: 6-28.

Bahl, P. and V.N. Padmanabhan, 2000. RADAR: An in-building RF-based user location and tracking system. Proceedings of 19th Annual Joint Conference of the IEEE Computer and Communications Societies, (INFOCOM 2000), Tel Aviv, Israel, pp: 775-784.

Bonivento, A., C. Fischione, L. Necchi, F. Pianegiani and A. Sangiovanni-Vincentelli, 2007. System level design for clustered wireless sensor networks. IEEE Trans. Industrial Inform., 3: 202-214.

Chong, L., K. Wu and J. Pei, 2007. An energy-efficient data collection framework for wireless sensor networks by exploiting spatiotemporal correlation. IEEE Trans. Parallel Distrib. Syst., 18: 1010-1023.

Heinzelman, W.B., A.P. Chandrakasan and H. Balakrishnan, 2002. An application-specific protocol architecture for wireless microsensor networks. IEEE Trans. Wireless Commun., 1: 660-670.

Lindsey, S. and C.S. Raghavendra, 2002. PEGASIS: Power-efficient gathering in sensor information systems. Proceedings of IEEE Aerospace Conference, 2002 Los Angeles, CA, USA, pp: 1125-1130.

Lindsey, S., C.S. Raghavendra and K.M. Sivalingam, 2002. Data gathering algorithms in sensor networks using energy metrics. IEEE Trans. Parallel Distrib. Syst., 13: 924-935.

Ming, Y., K.K. Leung and A. Malvankar, 2007. A dynamic clustering and energy efficient routing technique for sensor networks. IEEE Trans. Wireless Commun., 6: 3069-3079.

Muruganathan, S.D., D.C.F. Ma, R.I. Bhasin and A.O. Fapojuwo, 2005. A centralized energy-efficient routing protocol for wireless sensor networks. IEEE Commun. Maga., 43: S8-13.

Sangho, Y., P. Geunyoung, H. Junyoung, H. Jiman and J. Gwangil *et al*., 2005. Energy aware routing based on adaptive clustering mechanism for wireless sensor networks. Proceedings of the 3rd International Conference on Networking and Mobile Computing, (ICCNMC 2005), Zhangjiajie, China, pp: 1115-1124.

Seung, J.B., V. Gustavo de and S. Xun, 2004. Minimizing energy consumption in large-scale sensor networks through distributed data compression and hierarchical aggregation. IEEE J. Sel. Areas Commun., 22: 1130-1140.

Sevgi, C. and A. Kocyigit, 2008. On determining cluster size of randomly deployed heterogeneous WSNs. IEEE Commun. Lett., 12: 232-234.

Younis, O. and S. Fahmy, 2004. HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE Trans. Mobile Comput., 3: 366-379.

Younis, O., M. Krunz and S. Ramasubramanian, 2006. Node clustering in wireless sensor networks: Recent developments and deployment challenges. IEEE Network, 20: 20-25.

Zhi, Q., A. Subramanian and A.H. Sayed, 2007. REACA: An efficient protocol architecture for large scale sensor networks. IEEE Trans. Wireless Commun., 6: 3846-3855.