An energy-aware multilevel clustering algorithm for wireless sensor networks

Xinfang Yan
Zhengzhou University, iexfyan@zzu.edu.cn

Jiangtao Xi
University of Wollongong, jiangtao@uow.edu.au

Joe F. Chicharo
University of Wollongong, chicharo@uow.edu.au

Yanguang Yu
University of Wollongong, yanguang@uow.edu.au

Follow this and additional works at: https://ro.uow.edu.au/infopapers

Part of the Physical Sciences and Mathematics Commons

Recommended Citation
Yan, Xinfang; Xi, Jiangtao; Chicharo, Joe F.; and Yu, Yanguang: An energy-aware multilevel clustering algorithm for wireless sensor networks 2008, 387-392.
https://ro.uow.edu.au/infopapers/1582

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au
An energy-aware multilevel clustering algorithm for wireless sensor networks

Abstract
Clustering sensors nodes as the basic of routing is an efficient mechanism for prolonging the lifetime of wireless sensor networks. In this paper, the high-efficient multilevel clustering is abstracted as a root tree which has the performances of the minimal relay set and the maximal weight according to graph theory. A mathematical model for the clustering virtual backbone is built. Based on the model, an algorithm called energy-aware multilevel clustering (EAMC) is proposed. The EAMC can reduce the number of relays used for data transmission by minimizing the amount of the nodes in the root tree (that is cluster-head). Furthermore, the algorithm combining with periodical rotation of cluster-heads, enable the energy load be evenly distributed among all the sensors in the network. The maintenance and update for EAMC algorithm is also presented in this paper. Simulation results demonstrate that the proposed approaches significantly increase the lifetime of the sensor network.

Keywords
energy, aware, multilevel, clustering, algorithm, for, wireless, sensor, networks

Disciplines
Physical Sciences and Mathematics

Publication Details

This conference paper is available at Research Online: https://ro.uow.edu.au/infopapers/1582
An Energy-Aware Multilevel Clustering Algorithm for Wireless Sensor Networks

Xinfang Yan
College of Information Engineering
Zhengzhou University
Zhengzhou 450052, China
Email: iexfy@zzu.edu.cn

Jiangtao Xi, Joe F. Chicharo and Yanguang Yu
School of Electrical, Computer and Telecommunications Engineering
The University of Wollongong
NSW 2522, Australia

Abstract—Clustering sensors nodes as the basic of routing is an efficient mechanism for prolonging the lifetime of wireless sensor networks. In this paper, the high-efficient multilevel clustering is abstracted as a root tree which has the performances of the minimal relay set and the maximal weight according to graph theory. A mathematical model for the clustering virtual backbone is built. Based on the model, an algorithm called Energy-Aware Multilevel Clustering (EAMC) is proposed. The EAMC can reduce the number of relays used for data transmission by minimizing the amount of the nodes in the root tree (that is cluster-head). Furthermore, the algorithm combining with periodical rotation of cluster-heads, enable the energy load be evenly distributed among all the sensors in the network. The maintenance and update for EAMC algorithm is also presented in this paper. Simulation results demonstrate that the proposed approaches significantly increase the lifetime of the sensor network.

Keywords- wireless sensor networks; multilevel clustering; root tree; minimal relay set

I. INTRODUCTION

The wireless sensor network (WSN) is an innovative technology for information acquisition and processing which integrates sensor technology, Micro Electromechanical Systems (MEMS) and network communications. WSN has been an active research field due to the wide application areas including military, industry, medical treatment, transportation, among others [1]. Despite the extensive work carried out achievements made in this area, there is still a few key research issues that are outstanding and in need of solutions. As the wireless nodes in a WSN are usually driven by power sources (e.g., batteries) which are not replaceable, the lifetime of a WSN largely depends on the life of the power sources. Hence improvement of the efficiency of energy consumption in a WSN is a crucial issue. In order to achieve this, routing methods based on clustering are commonly used [2].

Low-energy adaptive clustering hierarchy (LEACH) [3] is first proposed as clustering routing protocol in WSN. With LEACH the nodes in a WSN are divided into clusters, each containing a few nodes that are close to each other. In every individual cluster a node is selected as the head which acts as the data control centre and are responsible for local data fusion for the associated cluster. These cluster heads then construct a virtual backbone network (VBN) for data transmission among different clusters and the base station. Obviously, use of LEACH results in significant reduction in the amount of data required to be transmitted and the number of relays for data transmission, and accordingly the energy dissipation can also be significantly reduced. However, LEACH also suffers from a few shortcomings. Firstly, the cluster heads are selected in a random manner, and hence head selection usually is not even among all the nodes within a cluster, resulting in uneven power dissipation among these nodes. Secondly, cluster formation is required after every data transmission, involving excessive computation and control and hence energy dissipation. Thirdly, cluster heads directly communicate with the base station, which may also incur high energy dissipation due to their long distance to the base station.

In order to further improve the performance of LEACH, a number of approaches are proposed to modify LEACH technique in recent years, for example, EECS [4], LEACH-B [5], etc. With EECS, the cluster-head is chosen based on the distance to the base station (BS) and the node with the shortest distance is selected as the cluster-head. EECS can only balance energy among cluster-heads and it can not balance the energy in whole network. In [6], the cluster-heads are selected based on the residual energy and the one with most residual energy is chosen. The main disadvantage of this algorithm is the requirement of the energy information of all nodes of the network. HEED (hybrid energy-efficient distributed clustering) [7] periodically selects cluster heads according to the node residual energy and intra-cluster communication cost. HEED is characterized by low message overhead, and fairly uniform cluster-head distribution across the network. However, all these algorithms [4-7] include LEACH are of single-hop in nature, where all cluster-heads directly communicate with the BS. The main disadvantage of these single-hop clustering techniques is that the energy dissipation can be very high for long distance transmission between cluster-heads and BS, as the energy dissipation is directly proportional to the exponent of transmission distance.

Multi-hop clustering [8-14] are proposed with the aim to further improve energy conservation. With multi-hop clustering the cluster-heads are connected as chains, and cluster heads on the chain are employed to relay data to the BS. This multi-clustering mechanism is able to balance the energy...
consumption among all sensor nodes and achieves an obvious improvement on the network lifetime. However there are also problems with existing multi-hop clustering schemes. For example, PEGASIS in [8] requires the topology information of whole network. In [9], cluster-head selection is still on random basis. A promising approach is that the minimum connected dominating set (MCDS) is used as a high-level virtual backbone network [10-13]. The idea is as follows. By expressing all the nodes as a set V, a dominating set (DS) D of G is a subset of V such that any node not in D has at least one neighbour in D. If the induced subgraph G[D] of D is connected, then D is a connected dominating set (CDS). The method is to form a CDS with least number of cluster-heads and to properly coordinate all the other nodes in the network. However these techniques in [10-13] are all designed for ad hoc networks which are not suitable for WSN, in which energy saving is a major issue.

EAHC proposed in [14] is an energy-aware hierarchical clustering algorithm with lower message complexity and time complexity for WSN. The method guarantees that nodes with highest residual energy and lowest communication cost are selected as the cluster-heads. Also cluster-heads are periodically rotated and hence the energy load can be evenly distributed among all the sensors in the WSN. A shortcoming of EACH is that it does not consider the issue of minimizing the number of cluster heads.

In this paper, we present to employ a spanning root tree with minimal relay set with maximal weight [15] as the MCDS for the virtual communication infrastructure of WSN. A mathematical model for the virtual backbone is produced and under the proposed model, we propose a Energy-Aware MultiLevel Clustering algorithm (EAMC). The proposed EAMC has the following salient features in contrast to the work in [14]. Firstly, the number of cluster heads is decreased by designing the second metric. Also, when the clustering tree is formed, the amount of relay nodes is minimized in order to further reduce the amount of data transmission among the tree nodes. In addition, the proposed technique is adaptive in that sensor nodes can be automatically added to or removed from the WSN, resulting in improvement in terms of the robustness of the virtual infrastructure.

This paper is organized as follows. Section II gives the model and problem description. Section III proposes the EAMC algorithm. Simulation results are presented in Section IV, and Section V concludes the paper.

II. MODEL AND PROBLEM DESCRIPTION

A. The Network Model

In this paper we utilize a connected, undirected and weighted graph to describe a WSN with n nodes. The graph is denoted as $G = (V, E, w_1, w_2)$, where $V = \{v_0, v_1, \ldots, v_q\}$ is a set of vertices, representing the wireless sensor nodes, where $v_0$ is defined as the base station BS. The connections between the sensor nodes are given by $E = \{e_1, \ldots, e_m\}$, referred to as the edge set and edge $e_i \{v_j, v_k\}$ implies that nodes $v_j$ and $v_k$ are within the radio scope of each other. Each wireless node is characterized by two weight parameters. $w_1(v_i)$ is the first weight of vertex $v_i$, representing residual energy of the sensor node, and $w_2(v_i)$ is the second weight of vertex $v_i$, standing for the number of its neighbours.

B. The Energy Model for Communications

We use a simple model of [3] shown in Fig. 1 for the radio hardware energy dissipation. We assume that the transmitter and receiver consume same amount of energy $E_{elec}$ for transmitting and receiving one bit of message, and the transmit amplifier requires energy $E_{amp}$ for transmitting one bit of message over one meter (note that the unit is J/bit/m²). We also assume a $d^2$ energy loss due to channel transmission. Thus, to transmit a k-bit message in a distance $d$ between the transmitter and the receiver, the radio energy expends:

$$E_{tx}(k, d) = E_{elec} \times k + E_{amp} \times k \times d^2$$

(1)

and to receive this message, the radio energy expends:

$$E_{rx}(k) = E_{elec} \times k$$

(2)

C. The Mathematical Model for virtual backbone

Let us consider a WSN consisting of $n$ sensor nodes deployed over a square sensing field. A BS is located outside the sensing field. For the purpose of energy saving, we will construct a virtual backbone network with $q$ cluster-head nodes. Note that virtual backbone network must cover the entire WSN, but $q$ should be as small as possible. The above problem can be abstracted as the formation of a root tree with minimal relay set and with maximal weights ($w_1(v_i)$ and $w_2(v_i)$). Fig. 2 shows the relationships between the root tree and multilevel clustering.

![Figure 1. Radio Energy Dissipation Model](image)

Based on the above model we can see that we should not only minimize the transmission distance, but also reduce the number of transmit message operations in the algorithm design.

![Figure 2. root tree vs. clustering of sensor networks](image)
DFS (Depth First Search) algorithm.

The relay set $R$ is defined as:

$$y_k = \begin{cases} 1, v_k \in R \\ 0, v_k \notin R \end{cases}, \quad a_{pq} = \begin{cases} 1, v_p \in N_1(v_q) \\ 0, v_p \notin N_1(v_q) \end{cases}$$

where, $k = 1, 2, \ldots, n$; $v_k \in R$; $p = 1, 2, \ldots, m$; $v_p \in V$; $q = 1, 2, \ldots, n$; $N_1(v)$ is the one-hop neighbor set of node $v$.

For $\forall v \in V$, $R$ is the relay set of a root tree means that:

$$\sum_{i=1}^{m} a_{ik} \cdot y_k \geq 1, \quad k = 1, 2, \ldots, n$$

Based on above, the mathematical model of multilevel clustering can be represented with a programming function of multi-objective and multi-constraint.

Objectives: $\min \sum_{i=1}^{m} y_i$

$$\max \sum_{i=1}^{m} w(v_i) \cdot y_i, \quad w(v) = w_1(v) \text{ or } w_2(v)$$

Constraints: $\sum_{i=1}^{m} a_{ik} \cdot y_k \geq 1, \quad k = 1, 2, \ldots, n$

$$a_{ik}, y_i = 0 \text{ or } 1$$

Finding the minimum relay set is the problem of determining the minimum connected dominating set (MCDS) in graph theory, which is then a non-deterministic polynomial complete (NP-C) problem. This problem can only be solved using approximation algorithm for particular practical applications. We propose a simple and efficient heuristic distributed EAMC which adopts combination of Greedy and DFS (Depth First Search) algorithm.

III. THE EAMC CONSTRUCTION ALGORITHM

A. The EAMC Construction Algorithm

We still consider that the WSN comprises a BS and a set of sensors with each sensor having a unique ID label, and all sensor nodes except for the BS are static and power constrained with the same initial energy $E_0$. Also we assume that each individual sensor node has the knowledge of its energy level and it is able to control its transmission power. The EAMC algorithm aims to build an energy efficient root tree as the virtual backbone network for the entire network. The EAMC tree roots from the BS and spans the entire sensor network.

In order to clarify our analysis, let us define the following variables or parameters:

- $hop\_count(u, v)$ – the number of edges on the shortest path from $u$ to $v$, for $\forall u, v \in V$.
- $N_i(v) = \{ u \mid hop\_count(u, v) \leq 1 \}$ – the one-hop (open), the direct neighbour set of $v \in V$.
- $unexplored(v)$ – a set including the neighbours of $v$, which are waiting for being driven to run the EAMC.
- $parent(v)$ – a set for the parent of $v$ in the root tree.
- $children(v)$ - a set for the children of $v$ in the root tree.
- $l(v)$ – the location of node $v$ in terms of the level or the height with respect to the root tree.
- $state(v, l(v), parent(v), x)$ – a state after $v$ run EAMC, where $x$ is the next node who will run the EAMC.
- $from(v)$ – the node who has driven $v$ to run EAMC.

The algorithm starts from initialization where for all $v$ the above parameters are set as follows $parent(v):=\emptyset$, $children(v):=\emptyset$, $N(v):=\emptyset$ and $from(v):=\emptyset$. Then all nodes broadcast their own ID.

When receiving an ID message node $v$ records it into its $unexplored(v)$ and set $N(v):=N(v)+1$, and then broadcasts its two weights $w_1(v)=E(v)$ and $w_2(v)=N(v)$.

When receiving the weight information from all the neighbours, the nodes in $unexplored(v)$ will be reordered according to the rank of weights from high to low. This can avoid repeat compare and hence save runtime and energy costs.

Following the initialization described above the spanning root tree is generated according to the follow procedure:

1) The procedure starts from the BS, which sets itself as the tree root with the parameters $(l(BS):=0; parent(BS):=\emptyset)$. The BS then chooses the first node $x$ from its $unexplored(BS)$ (which has the highest residual energy or is with the largest neighbour number when the residual energy is equal) as the next node to run the algorithm. Meanwhile, the BS broadcasts $state(v, l(v), parent(v), x) := state(BS, 0, \emptyset, v)$ to its neighbors.

2) When a node $u$ receives $state(v, l(v), parent(v), x)$ from $v$, its state will be determined according to Rule (A) when $u = x$, otherwise the state will be determined according to Rule (B). These two rules are described as follows:

Rule(A): Node $u$ chooses a neighbour node $y$ with the smallest level (or the largest residual energy when the levels are equal) as its parent and set $parent(u):=y$, and this will minimize the height (level) of the tree. Then the parameters will be updated as $l(u):=l(v)+1$; from($u$):=v; unexplored($u$):=unexplored($u$)-{v}; If unexplored ($u$)$=\emptyset$, then chooses the first node $x$ from its $unexplored(u)$ as the next node to be explored. Meanwhile it sends $state(u, l(u), y, x)$ to its neighbours. However, if unexplored($u$)$=\emptyset$, implying that all neighbours are explored, it will broadcast a $state(u, l(u), y, v)$ to its neighbors and exit the algorithm.

Rule(B): The node $u$ sets $unexplored(u):=unexplored(u)-\{v\}$ and records $l(v)$: if node $v$ is child of $u$, node $u$ will record $v$ in its $children(u)$ and then wait for being explored.

The EAMC algorithm is terminated when the BS receives the state message from all of its neighbours. The nodes in the set of different $parent$ are the cluster-heads of each level. The
cluster numbers of each cluster-head is the nodes in its children set.

B. Performance Analysis of EAMC algorithm

The message complexity and the time complexity are two key performances for measuring multilevel clustering in WSN. The message complexity is defined as the total number of messages transmitted during the operation of the algorithm, and the time complexity is defined as the total run time of all the nodes during the operation of the algorithm. These two metrics can be evaluated as follows.

Assuming that the WSN contains \( n \) nodes. The initialization phase requires for every node to firstly send the ID, and then another message for the weights. Hence the message complexity is \( O(n) \). In the phase of constructing EAMC, each cluster-head sends a message to the nodes in its unexplored area after its state is confirmed. In the worst case, that is, all nodes have \( \Delta \) neighbours, the total number of messages is \( \Delta c \). Therefore, the maximal message complexity of EAMC is \( \max\{O(n), O(\Delta c)\} \), where \( \Delta \) is the degree of nodes, \( c \) is the number of cluster-heads. Because the algorithm is based on DFS, the time complexity is the operating time throughout all nodes in the WSN, which is \( 2n \). Therefore the total time complexity is \( O(n) \).

From the analysis above, the message complexity and time complexity of EAMC algorithm are all far lower than LEACH’s [3] (at least \( O(n^2) \)), especially in terms of message complexity which has vital impact on the lifetime of the WSN. On the other hand, the EAMC algorithm chooses clusterheads based on two metrics \( w_1(v) = E(v) \) and \( w_2(v) = N_v \). The first metric guarantees that the nodes with larger residual energy \( E(v) \) have high probability to become clusterheads. As \( E(v) \) are equal for all the nodes in the beginning of the process, the second metric is very important to assure that the nodes with larger \( N_v \) are given higher priority as the tree nodes. This can minimize the height of the tree. Only if these two weights are same we consider ID as a metric. We also update traditional DFS approach through optimization rule (a) to decrease the level of the tree and the amount of clusterheads, so that the energy cost for relaying messages is reduced.

C. Maintenance and update of the EAMC

Sensor nodes are prone to failure and may shutdown without any notifications in advance because of various reasons. New sensor nodes can be deployed at any time as well. In order to tackle these problems we propose a high efficient maintenance and update algorithm of the EAMC, which can adapt to these changes automatically when the sensor nodes are added and removed at any time.

1) Addition of new nodes

When a new sensor node \( v \) is deployed, it will firstly broadcast a joining request. Based on the response the new node \( v \) will update its neighbour set \( N_v \). If cluster-head nodes are found in its \( N_v \), the node \( v \)'s state will be decided according to Rule (a). Otherwise, if no any cluster-head node is in its \( N_v \), \( v \) can decide its state according to Rule (b). These two rules are as follows:

- **Rule(a):** Node \( v \) choose the cluster-head node \( u \) with lowest level as its parent and set \( l(v) := l(u) + 1 \); it will send \( parent(v) := u \) to node \( u \) and then exits the algorithm. Accordingly, node \( u \) sets \( children(u) := children(u) \cup \{v\} \) after receiving \( parent(v) := u \) and then exits the algorithm.

- **Rule(b):** If \( N_v \) does not include a cluster head, the node \( w \) with lowest level in its \( N_v \) is chosen as its parent and set \( l(v) := l(w) + 1 \); it also sends \( parent(v) := w \). Upon receiving \( parent(v) := w \), node \( w \) will sets itself as a cluster head and sets \( children(w) := \{v\} \). Then it exits the algorithm.

2) Removal of existing nodes

When a leaf node \( v \) breaks down, only its parent \( u \) deletes it from \( children(u) \) and sets \( children(u) := children(u) \setminus \{v\} \). If a tree node becomes off line, a message will be sent by its neighbour tree nodes along other tree branches to the BS and the BS will rerun the EAMC.

3) Examples

Assuming that 20 nodes with different weights are randomly distributed in a field of 100m×100m.

![Examples for addition and removal of nodes](image)

Figure 3. Examples for addition and removal of nodes
Fig. 3(a) shows the network topology of 20 nodes when the distance for direct transmission, denoted as \( d \), is 40 m. Fig. 3(b) shows the structure of multilevel clustering root tree after the EAMC is run. Fig. 3(c) shows two new nodes (21 and 22) are joined. Fig. 2(d) shows the structure of multilevel clustering root tree after node 21 and node 22 are joined. Fig. 3(e) and (f) show respectively the structure of multilevel clustering root tree after node 14 and node 16 are break down.

**IV. SIMULATION RESULTS AND ANALYSIS**

We run simulations to evaluate the performance of this algorithm. Assuming that 100 nodes are randomly distributed in a field of 100 m \( \times \) 100 m. Fig. 4 shows the network topology of 100 nodes when \( d = 40 \).

Each node is initially given 0.5 J of energy. We assume that the radio dissipates \( E_{\text{elec}} = 50 \text{nJ/bit} \) to run the transmitter and receiver circuitry and \( \varepsilon_{\text{amp}} = 100 \text{ pJ/bit/m}^2 \) for the transmit amplifier to transmit a bit of information for one meter. The length of the message packet is 128 bits.

Fig. 5 shows the structures of multilevel clustering root tree after the EAHC and EAMC are run. The amount of cluster heads in the EAMC(12) is fewer than ones in the EAHC(17). And the distribution of cluster heads in the EAMC is more evenly.

Fig. 6 illustrates number of cluster heads varies as the communication radius, \( d \), increases. If the communication radius of the node is too small, the network would be unconnected. Based on network connectivity (15-95 m), it is clearly seen that, the number of cluster heads in both EAHC and EAMC decreases gradually as the communication radius increases. The reason is that the number of neighbors of each node will decrease as the communication radius increases.

Fig. 7 illustrates how the system lifetime varies as \( d \) increases. Note that the lifetime is defined as the number of rounds for data collections by the BS before the first node dies. It is clearly seen that the lifetime decreases as the communication radius increases. The main reason is that the energy dissipation for date transmission among cluster heads increases because the energy dissipation is direct proportion of the transmission distance \( d^2 \) and \( d^4 \). The results show that we
should limit the transmission power to a small level in order to have a long system lifetime.

Figure 8. Lifetime vs. clustering round

Figure 8 shows the system lifetime using EAMC, utilizing periodic rotation of cluster-heads. The x-axial represents the rounds of data collections between two successive rotations of the cluster heads. It can be seen that, when the number of the rounds increases from 0 to 100, the lifetime increases sharply. The lifetime remains relatively stable on a high-value when the number of the increases from 100 to 1500. However, when the number of the rounds between two successive rotations exceeds 2500, there is a significant reduction in the lifetime. These results show that, the number of rounds between two successive rotations should be kept within the range from 100 to 1500 in order for EAMC having a long system lifetime.

V. CONCLUSION

We have presented a new approach for building a virtual backbone as the communication infrastructure of a WSN. The proposed approach is based on the formation of a root tree with minimal relay set while its nodes are selected to have maximal residual energy (or with the largest neighbour number when the residual energy is equal). In order to achieve this, a new energy-efficient adaptive clustering algorithm, referred to as EAMC, has been proposed, which is a distributed algorithm, and each individual nodes only require knowledge of their one-hop neighbours rather than the knowledge of the whole network. The proposed technique has small message complexity and time complexity. Comparing with EAHC, the EAMC has fewer cluster heads by introducing a few optimization rules and thus further reduce the amount of data delay transmission among the tree nodes in the network, resulting in a longer network lifetime. The paper also proposed a scheme for the maintenance and update of the network. The proposed scheme is able to improve the robustness of the communications infrastructure in the cases when new sensor nodes are deployed or some existing sensor nodes are removed from the network. Simulation results have shown the correctness and the ability of self-renew.

ACKNOWLEDGEMENT

This work was partially supported by the Natural Science Foundation of Henan Province of China under Grant No. 72300410430.

REFERENCES

[15] J.H. Yin and K.Y. Wu, Graph Theory with Algorithm, China Science and Technology Press, HeFei, 2004