

## A study on clustering attributes in wireless networks

<sup>1</sup> Aravind Kumar Sharma, <sup>2</sup> Nitesh Kumar

<sup>1</sup> Assistant Professor, Department of Computer Science, OPJS University Churu, Rajasthan, India

<sup>2</sup> Research Scholar, OPJS University Churu, Rajasthan, India

### Abstract

Hybrid Energy-Efficient Distributed clustering (HEED) introduced by Younis and Fahmy, is a multi-hop WSN clustering algorithm which brings an energy-efficient clustering routing with unambiguous thought of energy. Different from LEACH in the manner of CH election, HEED does not select nodes as CHs arbitrarily.

In HEED, CHs are periodically elected based on two important parameters: residual energy and intra-cluster communication cost of the candidate nodes. Initially, in HEED, a percentage of CHs among all nodes, Cprob, is set to assume that an optimal percentage cannot be computed a priori. The probability that a node becomes a CH is:  $P = \frac{E_{residual}}{E_{max}}$  Where  $E_{residual}$  is the estimated current energy of the node, and  $E_{max}$  is a reference maximum energy, which is typically identical for all nodes in the network. The value of Cprob, however, is not allowed to fall below a certain threshold that is selected to be inversely proportional to  $E_{max}$ . Afterwards, each node goes through several iterations until it finds the CH.

**Keywords:** wireless, network, node

### Introduction

LEACH (low-energy adaptive clustering hierarchy) is one of the pioneering clustering routing approaches for WSNs. It divides the protocol operation into rounds, and each round is subdivided into two phases: setup and steady-state phase. In the setup phase, the nodes create clusters and elect CH. The node becomes a CH for the current round if the number is less than the following threshold:

In the case of the steady state phase, non-CHs transmit the sensed data to their CH. CHs receive the data, aggregate it into a single packet and forward it to the BS. After a certain period

of time the network returns to the setup phase. However, LEACH has some drawbacks-

- i) It does not take account of the energy level of the sensor nodes.
- ii) It does not employ any scheme to control the cluster.
- iii) Additionally, the cluster can be formed in a disproportional way, which causes furthermore communication interferences and network partitioning.
- iv) LEACH cannot provide reliability, energy-efficiency or a fair distribution of resources.

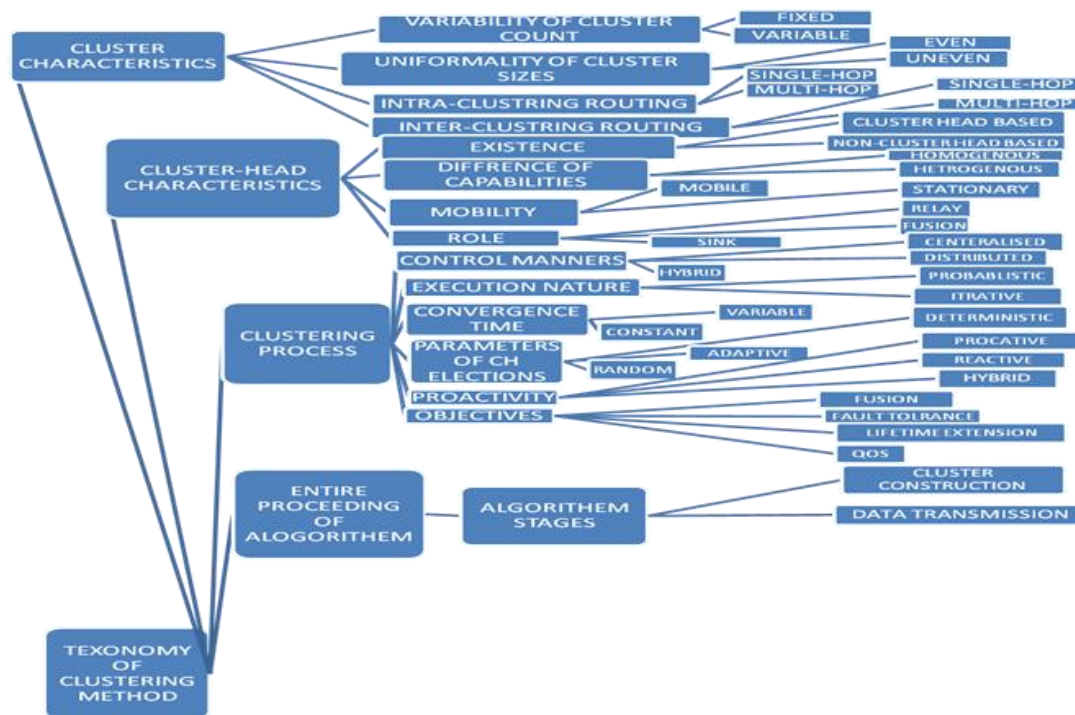


Fig 1: Taxonomy of Clustering Methods in WSNs

The advantages of the HEED protocol are as follows: (1) It is a fully distributed method that aids from the use of the two important parameters for CH election; (2) Low power levels of clusters promote an increase in spatial reuse while high power levels of clusters are required for inter-cluster communication. This provides uniform CH distribution across the network and load balancing; (3) Communications in a multi-hop fashion between CHs and the BS promote more energy conservation and scalability in contrast with the single-hop fashion, i.e., long-range communications directly from CHs to the sink, in the LEACH protocol.

However, there are some limitations with HEED as follows: (1) the use of tentative CHs that do not become final CHs leave some uncovered nodes. As per HEED implementation, these nodes are forced to become a CH and these forced CHs may be in range of other CHs or may not have any member associated with them. As a result, more CHs are generated than the expected number and this also accounts for unbalanced energy consumption in the network; (2) Similar to LEACH, the performing of clustering in each round imposes significant overhead in the network. This overhead causes noticeable energy dissipation which results in decreasing the network lifetime; (3) HEED suffers from a consequent overhead since it needs several iterations to form clusters. At each iteration, a lot of packets are broadcast. (4) Some CHs, especially near the sink, may die earlier because these CHs have more work load, and the hot spot will come into being in the network.

### **TI-LEACH**

Two-Level Hierarchy LEACH (TL-LEACH) is an addition to the algorithm of LEACH. TL-LEACH uses the following two techniques to achieve energy and latency efficiency: randomized, adaptive, self-configuring cluster formation and localized control for data transfers. A CH collects data from MNs as original LEACH, but instead of transmitting data to the BS directly, it uses a part of CHs that lies between the CH and the BS as a relay station.

The algorithm has four basic phases: advertisement phase, cluster setup phase, schedule creation and data transmission. In the first phase, each node decides whether it become a primary CH, secondary CH or ON in each round which is the same as that of LEACH. If a node is elected a primary CH, it must advertise other nodes. The mechanism used in this phase is carrier sense multiple access (CSMA). Thereafter, secondary CH nodes send the advertisement to the ONs. In this phase, each secondary CH decides to which primary CH it belongs and sends an advertisement message to its primary CH. Each ON must decide which secondary CH it belongs to and informs it through an opposite message. In the third phase, each primary CH creates a TDMA schedule assigning each node in its group a slot to transmit.

The advantages of TL-LEACH are as follows: (1) Uses random rotation of local cluster BSs; (2) TL-LEACH uses confined synchronization, which is conducive to scalability and robustness in the network; (3) Compared with LEACH, the scheme of two-levels clustering leads to less average transmission distance, and less nodes are required to transmit far distances to the BS via TL-LEACH. This effectively reduces the total energy consumption.

Disadvantages of TL-LEACH are: (1) Despite that the average communication distance is decreased in comparison with

LEACH, the two-hop inter-cluster routing of TL-LEACH is still not applicable to large-range networks.

### **EECS**

Energy Efficient Clustering Scheme (EECS) is a clustering algorithm which better suits the periodical data gathering applications. EECS is a LEACH-like scheme, where the network is partitioned into several clusters and single-hop communication between the CH and the BS is performed. In EECS, CH candidates compete for the ability to elevate to CH for a given round. This competition involves candidates broadcasting their residual energy to neighboring candidates. If a given node does not find a node with more residual energy, it becomes a CH. Different from LEACH for cluster formation, EECS extends LEACH by dynamic sizing of clusters based on cluster distance from the BS.

The advantages of EECS are as follows: (1) EECS constructs balancing point between intra-cluster energy consumption and inter-cluster communication load based on energy and distance; (2) Clustering is performed by dynamic sizing based on cluster distance from the BS.

However, there exist a few disadvantages in EECS as follows:

- 1) Single-hop communications account in EECS, long-range transmissions directly from CHs to the BS can lead to much energy consumption. So, not suitable for large-range networks.
- 2) More global knowledge is required about the distances between the CHs and the BS, and the task of global data aggregation adds overheads to all sensor nodes.

### **EEUC**

Energy-Efficient Uneven Clustering (EEUC) algorithm is a distributed competitive and clustering algorithm, where CHs are elected by confined competition, which is not like LEACH. Each node has a pre-assigned viable range, which is smaller as it gets close to the BS.

This gives EEUC an uneven clustering approach for the determination of balancing energy consumption among CHs. In the process of CH election, each node generates a random number, and only the node whose number is greater than a threshold will be activated for CH election by broadcasting compete message within a competition radius which is determined by its distance to the BS.

The disadvantages of EEUC are as follows:

- 1) To report the hot spots problem, EEUC introduces an uneven clustering mechanism to balance the energy consumption among CHs. Accordingly; the uneven clustering mechanism in EEUC improves the network lifetime over LEACH and HEED.
- 2) Based on communication cost, this protocol can save more energy via inter-cluster multi-hop routing mechanism in steady state phase; because a CH would choose its relay node from the two whose communication cost are the least among all of its neighbor CHs.

### **ACE**

Algorithm for Cluster Establishment (ACE) employs an emergent algorithm, which is any calculation that achieves formally or stochastically probable global effects, by communicating directly with only a bounded number of instant neighbors and without the use of central control or global visibility. One of the main individual characteristics of

growing protocols over other localized protocols is the survival of feedback during protocol operation.

The point of ACE is to allow a node to assess its possible as a CH before becoming one and retire if it is not the best CH at the moment. The algorithm works in iterations that do not have to be synchronized at the individual nodes. ACE has two logical parts: the spawning of new clusters and the migration of existing clusters. When a node decides to become a CH, it spawns of new cluster by broadcasting an invitation message to recruit its neighbors. Upon getting the invitation, a neighboring node joins the new cluster and becomes a follower of the new cluster. At any moment, a node can be a follower of multiple clusters while the protocol is running. However, the node can be a loyal follower, i.e., a member which belongs to only one cluster. Migration is a process in which the best candidate for being CH is elected again. Each CH will occasionally check the ability of its followers to define which is the best candidate for the new leader of the cluster. The CH will retire if one of these followers has more followers than it does. The node, considered as the best candidate for CH, would have the largest number of followers while minimizing the amount of overlap with existing clusters. Once the best candidate is determined by the current CH, it will assign the best candidate as the new CH and abdicate its position as the old CH. Thus, the position of the cluster will appear to migrate in the direction of the new CH. Accordingly, some of the former followers of the old CH do not belong to the new cluster, while some new nodes near the new CH become new followers of the new cluster.

The characteristics and advantages of ACE are summarized as follows:

1. ACE is a growing algorithm that uses response to induce the formation of a highly efficient cover of uniform clusters over the network;
2. Minimizing the number of CHs would not only generate efficient cover of the whole network but also minimizes the cluster overlaps. This also improves the efficiency of the algorithms that executes at the level of the CHs;
3. ACE is very strong because it can easily repair structure damage caused by node failure and can also mix new nodes in the network.

### **BCDCP**

Base-Station Controlled Dynamic Clustering Protocol (BCDCP) is a centralized clustering routing protocol with the BS capable of difficult computation. The main idea of BCDCP is the cluster formation where each CH serves an almost equal number of MNs to balance CH overload and uniform CH placement throughout the network.

A few disadvantages in BCDCP as follows:

1. BCDCP is a centralized algorithm which brings poorer scalability and robust to large networks than distributed algorithms;
2. Each node needs to transmit information regarding its location and energy level to the BS during the process of cluster formation;
3. Due to the single-hop routing scheme, it is not appropriate for long-distance communications, which result in much energy consumption.

### **Data-Transmission Based Clustering Routing Protocols PEGASIS**

Power-Efficient Gathering in Sensor Information Systems (PEGASIS) is an improvement of LEACH. The main idea of PEGASIS is for each node to only communicate with their close neighbors and take turns being the leader for transmission to the sink. In PEGASIS, the locations of nodes are random, and each sensor node has the ability of data detection, wireless communication, data fusion and positioning. Energy load is distributed evenly among the sensor nodes in the network.

In PEGASIS, the nodes are organized to form a chain, which can either be concentrated assigned by the sink and broadcast to all nodes or accomplished by the nodes themselves using a greedy algorithm. If the chain is formed by the nodes themselves, they can first get the location data of all nodes and locally compute the chain using the same greedy algorithm.

During the process of chain formation in PEGASIS, it is assumed that all nodes have global knowledge of the network and the greedy algorithm is employed. The chain construction is commenced from the furthest node from the sink and the closest neighbor to this node will be the next node on the chain. When a node on the chain dies, the chain will be reconstructed in the same manner to bypass the dead node.

The following is the advantages of PEGASIS:

This protocol is able to outperform LEACH for different network sizes and topologies, because it reduces the overhead of dynamic cluster formation in LEACH, and decreases the number of data transmission volume through the chain of data aggregation; (2) The energy load is dispersed uniformly in the network. To ensure that the fixed sensor node is not select as the leader and thus to prevent the subsequent early death of this sensor node, all sensor nodes act as the leader in turn.

### **TEEN**

Threshold sensitive Energy Efficient sensor Network protocol (TEEN) is a hierarchical protocol whose main goal is to cope with sudden changes in the sensed attributes such as temperature. The protocol combines the hierarchical technique in line with a data-centric approach. The nodes sense their environment continuously, but the energy consumption in this algorithm can potentially be much less than that in the proactive network, because data transmission is done less frequently.

In TEEN and two thresholds, hard threshold and soft threshold, are defined. The former threshold is a threshold value for the sensed attribute. It is the absolute value of the attribute beyond which, the node sensing this value must switch on its transmitter and report to its CH. The latter threshold is a small change in the value of the sensed attribute which triggers the node to switch on its transmitter and transmit.

In TEEN, a CH sends its members a hard threshold and a soft threshold. Thus the hard threshold tries to reduce data communications by allowing the nodes to transmit only when the sensed attribute is in the range of interest. The soft threshold further reduces data communications might have otherwise occurred when there is little or no change in the

sensed attribute. At the expense of increased energy consumption, a smaller value of the soft threshold generates more accurate information of the network, thus users can control the trade-off between energy efficiency and data accuracy by the parameters adjustment. Moreover, the soft threshold can be varied and the users can change the fresh parameters as required at every cluster change time.

According to above discussion, TEEN has the following advantages:

- 1) Based on the two thresholds, data transmission can be controlled commendably, i.e., only the sensitive data we demand can be transmitted, so that it reduces the energy transmission consumption and improves the effectiveness and usefulness of the receiving data;
- 2) TEEN is complement for reacting to large changes in the sensed attributes, which is suitable for reactive scenes and time critical applications.

However, there exist a few drawbacks in TEEN as follows:

- 1) It is not suitable for periodic reports applications since the user may not get any data at all if the values of the attributes may not reach the threshold.
- 2) There exist wasted time-slots and a possibility that the BS may not be able to distinguish dead nodes from alive ones, because only when the data arrive at the hard threshold and has a variant higher than the soft threshold did the sensors report the data to the BS
- 3) (3) If CHs are not in the communication range of each other the data may be lost, because information propagation is accomplished only by CHs.

### APTEEN

The Adaptive Threshold sensitive Energy Efficient sensor Network protocol (APTEEN) is an extension to TEEN and aims at both transmitting periodic data and reacting to time critical events. APTEEN, on the other hand, is a hybrid protocol that changes the periodicity or threshold values used in TEEN according to the requirement of users and the type of the application. APTEEN is based on a query system which allows three types of queries: historical, on-time, and persistent which can be used in a hybrid network. Moreover, QoS requirements are introduced for the on-time queries and minimum delay is achieved by a TDMA schedule with a special time slot assignment manner.

In APTEEN, CHs broadcast the following four parameters:

- 1) Attributes (A)—this is a set of physical parameters which the user is interested in obtaining data about;
- 2) Thresholds—this parameter consists of the hard threshold (HT) and soft threshold (ST). HT is a particular value of an attribute beyond which a node can be triggered to transmit data. ST is a small change in the value of an attribute which can trigger a node to transmit data again;
- 3) Schedule—this is a TDMA schedule, assigning a slot to each node;
- 4) Count time (CT)—it is the maximum time period between two successive reports sent by a node. It can be a multiple of the TDMA schedule length and it accounts for the proactive component.

The distinctive feature of APTEEN is to switch between proactive and reactive modes to transmit data. All nodes sense the environment continuously, but only those nodes which

sense a data value at or beyond the hard threshold permit transmitting. Once a node senses a value at or beyond the hard threshold, it transmits data. If a node does not send data for a time period equal to the count time, it must sense and transmit the data again. In APTEEN, each CH aggregates the data from the MNs within its cluster and transmits the aggregated data to the BS. During the process of data aggregation, it is assumed that the data received from the corresponding MNs are sufficiently correlated, thus it reduces a large amount of redundancy of the data to be transmitted to the BS. Moreover, a modified TDMA schedule is used to implement the hybrid network by assigning each node in the cluster a transmission slot. Additionally, APTEEN offers a lot of flexibility by allowing the user to set the CT interval and the threshold values for energy consumption can be controlled by changing the CT as well as the threshold values.

The characters and advantages of APTEEN include:

- 1) APTEEN combines both proactive policies, which is alike that of LEACH, and reactive policies, which is alike that of TEEN. Accordingly it is suitable in both proactive and reactive applications;
- 2) It embodies a lot of flexibility by setting the count-time interval, and the threshold values for the energy consumption can be adjusted by changing the count time as well as the threshold values.

The main disadvantages of APTEEN are as follows:

- 1) there exist additional complexity required to implement the threshold functions and the count time;
- 2) Actually, both TEEN and APTEEN share the same drawbacks of additional overhead and complexity of cluster construction in multiple levels, implementing threshold-based functions, and dealing with attribute-based naming of queries— APTEEN more than TEEN.

### TTDD

The Two-Tier Data Dissemination (TTDD) approach is a low-power protocol for efficient data delivery from multiple sources to multiple mobile sinks. It exploits a geographic routing based on grid of cells as the routing method. Instead of passively waiting for queries from sinks, sensor nodes can proactively establish a structure to set up forwarding information. Ultimately, the sensing field is figured as a set of grid points.

In TTDD, a source divides the field into a grid of cells and each cell is square. A source, at one crossing point of the grid, propagates data announcements to reach all the other crossings, called dissemination points, on the grid as shown in Figure. A source calculates the locations of its four neighboring dissemination points and sends a data announcement message to the four neighboring dissemination points using simple greedy geographical forwarding, i.e., it forwards the message to the neighbor node that has the smallest distance to the neighboring dissemination point. Similarly, the neighbor node continues forwarding the data announcement message till the message stops at a node that is closer to the dissemination point than all its neighbors. During this process, each intermediate node stores the source information and this process continues until the message stops at the border of the network. After this process, the grid structure is obtained.

The character and advantage of TTDD can be concluded as follows:

- It deals with the problems caused by multiple mobile sinks and sink moving in large-scale WSNs;
- Despite that it is effective in high mobility scenarios, the overhead to build and maintain the overlay is significant, especially in periodic reporting scenarios, which are more traffic intensive than event-based reporting. Therefore, TTDD is better suited to event-detecting WSNs with sporadic rather than continuous traffic.

### CCS

The Concentric Clustering Scheme (CCS) to reduce the energy consumption loopholes in PEGASIS. The main idea of CCS is to consider the location of the BS to enhance its performance and to prolong the lifetime of the network.

In CCS, the network is divided into a variety of concentric circular tracks which represent different clusters and each circular track is assigned with a level. The track nearest to the BS is assigned with level-1 and the level number increases with the increase of the distance to the BS. Thus, each node in the network is assigned with its own level. Besides, chains are constructed within the track as that in PEGASIS. One of the nodes on the chain at each level area is selected as a CH. A CH in level L is selected with node number obtained by calculating  $i \bmod ML$ , where ML represents the number of nodes that have the same level in  $i$  round. Data transmission in CCS is based on the process of PEGASIS protocol. After CH selection, each CH transmits the data of its own location to both the upper and lower level CH in one grade. In the process of the data transmission, all nodes in each level transmit the data to the nearest node from themselves along the chain. The node receives the data and fuses its own data and transmits these data to the next node. Therefore, the CH receives at most two data messages. Subsequently, the CH in each level transmits the data to the lower CH. At last, level 1 CH transmits these data to the BS.

Compared to PEGASIS, CCS embodies the advantaged as follows:

- The distance over which the data can be transmitted to the BS from the CH is reduced in CCS. Hence, a considerable amount of energy is saved on account of the reduction of transmission distance in CCS;
- The network is divided into a series of concentric clusters, and the reverse data flow from the BS is also reduced. Thus, considerable amount of energy also conserved during data transmission.

### HGMR

Hierarchical Geographic Multicast Routing (HGMR), proposed in <sup>[24]</sup> by Koutsonikolas *et al.*, is a location-based multicast protocol. This protocol seamlessly incorporates the key design concepts of the Geographic Multicast Routing (GMR) and Hierarchical Rendezvous Point Multicast (HRPM) protocols and optimizes them by providing forwarding energy efficiency as well as scalability to large-scale WSNs.

HGMR starts with a hierarchical decomposition of a multicast group into subgroup of manageable size by means of the key concept of mobile geographic hashing of HRPM. Within each subgroup, HGMR adopts the local multicast scheme of GMR to forward data packets along multiple branches of the multicast tree in one transmission. In HGMR, the multicast

group is divided into subgroups using the mobile geographic hashing idea: the deployment area is recursively partitioned into a number of  $d^2$  equal-sized square sub-domains called cells, where  $d$  is decomposition index depending on the encoding overhead constraints, and each cell comprises a manageably-sized subgroup of members. In each cell there is an Access Point (AP) responsible for all members in that cell, and all APs are managed by a Rendezvous Point (RP).

In order to join a hierarchically decomposed multicast group, a node generates the hashed location for the RP and sends a join message to that location. After receiving the value of decomposition index  $d$  from the RP, the node invokes the hash function with  $d$  and its location, to achieve the hashed location of the AP of the cell it belongs to. Consequently, the source builds an overlay tree, the Source-to-AP tree, whose vertices are active APs in a topology graph, and another overlay tree, the AP-to-Member tree is also built from the AP, considering each member as the vertex. When a source needs to transmit data packets, it utilizes the unicast-based forwarding strategy of HRPM to propagate data packets to each AP along the Source-to-AP overlay tree. In each cell, instead of constructing an AP-to-Member overlay tree, HGMR uses the cost over progress optimizing broadcast algorithm of GMR to select the next relay nodes at each hop.

By adjusting the value for the decomposition index  $d$ , the number of members an AP is responsible for does not increase too much. Hence, the use of GMR within each cell instead of the unicast-based forwarding strategy of HRPM contributes to reduce the number of transmissions. When routing to a hashed location (RP or AP), HGMR uses the face routing of HRPM, while when routing from an AP to a set of group members within a cell, it uses the multicast face routing of HRPM.

The main merits of HGMR can be summarized as follows:

- 1) The membership management of HGMR is very simple and easy without additional cost due to the geographic hashing algorithm;
- 2) According to the number of the nodes which play the different roles, the data transmission methods for different hierarchies in HGMR make the routing energy-efficient in a way;
- 3) HGMR is free of the scalability problem in that only manageable destinations exist in a cell.

However, there are a few drawbacks in HGMR as follows:

- 1) The simple network partition into a set of cells may lead to sub-optimal routing paths from the root node to multicast group members;
- 2) All transmissions are concentrated to APs. APs can be changed to another node by hash function, but it is too limited in a cell, and may bring on unbalanced energy consumption around APs;

### References

1. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E. Wireless sensor networks: A survey. *Comput. Netw.* 2012; 38:393-422.
2. Al-Karaki JN, Kamal AE. Routing techniques in wireless sensor networks: A survey. *IEEE Wirel. Commun.* 2014; 11:6-28.
3. Sheng B, Li Q, Mao W, W Jin. Outlier Detection in Sensor Networks, in proceedings of the Eighth ACM International Symposium on Mobile Ad Hoc Networking

- and Computing (MobiHoc). 2007, 219-228.
4. Braginsky D, Estrin D. Rumor Routing Algorithm for Sensor Networks. In Proceedings of the First ACM International Workshop on Wireless Sensor Networks and Applications (WSNA), Atlanta, GA, USA, 2010, 22-31.
  5. Buttyan L, Schaffer P. PANEL. Position-Based Aggregator Node Election in Wireless Sensor Networks. In Proceedings of the 4th IEEE International Conference on Mobile Ad-hoc and Sensor Systems Conference (MASS), Pisa, Italy, 2007; 1-9.
  6. Buttyan L, Schaffer P. PANEL: Position-based aggregator node election in wireless sensor networks. *Int. J. Distrib. Sens. Netw.* 2010; 1-16.
  7. Chan H, Perrig A. ACE: An Emergent Algorithm for Highly Uniform Cluster Formation. In Proceedings of the 1st European Workshop on Sensor Networks (EWSN), Berlin, Germany, 19-21 January 2014; 154-171.
  8. Chao SHA, Ruchuan WANG, Hai ping HUANG, Li Juan SUN. Energy efficient clustering algorithm for data aggregation in wireless sensor networks. *The Journal of China Universities of Posts and Telecommunications* 17, Supplement 2010; 2(0):104-122.
  9. Das SM, Pucha H, Hu YC. Distributed hashing for scalable multicast in wireless ad hoc networks. *IEEE Trans. Parallel Distrib. Syst.* 2008; 19:347-362.
  10. Ding P, Holliday J, Celik A. Distributed Energy Efficient Hierarchical Clustering for Wireless Sensor Networks. In Proceedings of the 8th IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS), Marina Del Rey, CA, USA, 2005; 322-339.
  11. Dechene DJ, Jardali AE, Sauer MLA. A Survey of Clustering Algorithms for Wireless Sensor Networks, *Computer Communications*. Butterworth-Heinemann Newton, MA USA, 2007.
  12. Foster A, Murphy AL, FROMS: A failure tolerant and mobility enabled multicast routing paradigm with reinforcement learning for WSNs. *Ad Hoc Netw.* 2011; 9:940-965.
  13. Gautam N, Lee WI, Pyun JY. Track-Sector Clustering for Energy Efficient Routing in Wireless Sensor Networks. In Proceedings of the 9th IEEE International Conference on Computer and Information Technology (CIT), Xiamen, China. 2009: 116-121.