A SURVEY ON CLUSTERING TECHNIQUES FOR WIRELESS SENSOR NETWORKS

[1]Karishma Desai, [2]Keyur Rana
[1]Student, [2]Associate Professor
Email: [1]Khdesai91@gmail.com, [2]keyur.rana@scet.ac.in

Abstract— Wireless Sensor Network (WSN) includes set of sensors with sensing, communication and communications capability. Sensor nodes are energy constrained. They have only single source of energy and that is it battery. Once the sensor nodes have been deployed in the hostile environment it is not feasible to change or replace or restore battery. Therefore main challenge in WSN is efficient use of energy to increase network lifetime. In order to solve this problem various clustering techniques are introduced that organize the network into connected hierarchy. In clustering, sensor nodes are grouped into small clusters and a cluster head is elected for each cluster. For the large scale network, hierarchical model gathers the data at every cluster and sends the processed data towards the base station. This eliminates redundant data to be communicated, which in turn; helps to reduce energy.

Index Terms—Cluster, Cluster Head, Energy Efficiency, Wireless Sensor Networks

I. INTRODUCTION

Advances in wireless communication made it possible to develop network consisting of small size, low power and multifunctional sensor nodes. RF transceiver and batteries, CPUs, one or more microcontrollers for processing are embedded in the sensor nodes. The sensor nodes are deployed inside the network or very close to it. The sensor node can sense temperature, pressure, seismic, thermal, video, sound, soil [1], etc. and send information to the sink node. The sensor nodes have limited resources like limited battery, low memory, low bandwidth processing, etc. In spite of these limitations when sensor nodes are located in the network, they provide a real picture of the network being sensed. Once the nodes have been located it may possible that it will left unattended. In such scenario WSN can provide efficient way of communication.

Wireless sensor networks are widely used in various areas like commercial and industrial areas such as for e.g. target field imaging, weather monitoring, security and tactical surveillance, habitat monitoring, healthcare, process monitoring, detecting ambient conditions such as movement, sound, light, or the presence of certain objects [1,2, 3,5]. Many applications requires sensor network to operate for few months or years. In various applications once the nodes have been deployed in hostile environment it is very difficult to replace or recharge the battery. In such cases it becomes important to utilize energy in an efficient manner and extend network lifetime of the sensor network to provide services to such applications for the longer time.

A lot of energy is consumed in sensing activities in the network, processing the data, communicating with other nodes and transmitting the collected data to the base station (BS) using intermediate nodes. Network lifetime also depends on battery of sensor nodes. In order to extend network lifetime, it is necessary that energy of the nodes should be used efficiently.

The energy used in one bit of data transfer can be used to perform a large number of arithmetic operations in the sensor processor [4]. In a densely deployed sensor network the data sensed by the sensor will be similar and it will generate the same information. If we will transmit this data to the BS then redundant data will be transferred and more energy will be consumed. Therefore in order to transmit only unique information, it is advisable to arrange sensor nodes in some group. By arranging sensor nodes in the group, the data will be combined from all the sensor nodes and only compact data will be transmitted to the BS. This will reduce the data that need to be transmitted and it will consume less energy. The process of grouping of sensor nodes is known as clustering. In every cluster there will be one cluster head (CH). All sensor nodes will send their data to their respective CH and then CH will send the data to the BS. Clustering will reduce the communication overhead and decreasing energy consumptions and interferences among sensor nodes. Clustering provides various advantages like scalability, coverage, robustness, simplicity, load balancing, extended network lifetime, etc.

The remaining of this paper is organized as follows, Section II reviews of related work on different existing clustering scheme. Section III shows the comparison of different existing clustering methods. Section IV ends the paper with the conclusion.

II. RELATED WORK

Although there are many quality papers have been presented for clustering in WSNs. The goal of this study is to provide the survey of cluster based sensor network and to identify the research for future work.

W. Heinzelman et al. [9] proposed first dynamic clustering algorithm which makes use of randomized rotation of CHs among the cluster member to evenly balance the load of CH. This protocol helps in reducing amount of data that need to be transferred to the BS. This protocol is suitable for application where sensors don’t need to send continuous data. Each round is consists of two phases: setup phase and steady phase. In setup phase, clusters are being formed. Each sensor will generate one random number and this number is compared with threshold. If this number is less than the threshold then this sensor can become CH. Now which sensor is selected as CH will send broadcast message and informs other nodes in the network that it is CH. The nodes which are not CH receive the message and decide to which CH they should join based upon received signal strength of the message. Each non CHs node will send information to the CH that they are joining them.

In steady state phase data are transmitted. CH creates one TDMA schedule and passes this schedule to every cluster member. This schedule tells every cluster member when they should transmit the data to CH. Now cluster member will send their data during their allocated time slots. After collecting data from cluster member CH will perform data aggregation and aggregated data is passed to the BS.

Advantage of this algorithm is that cluster heads are not fixed in every round. Different set of CHs are selected for each round. It is energy efficient then the static method of clustering and direct diffusion. There are some limitations of this algorithm like it does not provide optimal number of CHs because it may possible that not a single node is selected as a CH during some round. Energy parameter is not considered while selecting the CH. It also may possible that all the selected CH will be on the same side of the network. There is no even distribution of the CHs. There is single hop clusters are formed which will consume more energy.

W. Heinzelman et al. [5], proposed a LEACH-Centralized (LEACH-C) algorithm. In order to overcome disadvantage of LEACH new improved centralized clustering algorithm was proposed. LEACH-C is a centralized algorithm in which CHs are decided by BS and clusters are also formed by the BS. In steady state phase, each node sends information about current energy level and position to the BS. BS decides one threshold level and the nodes that have energy higher than the threshold is selected as CH. After determining CH, BS will send a message containing CH. When node receives this message it compare CH ID with its own ID and if it matches then it is a CH else it is a cluster member. The problem of determining the optimal number of cluster heads is an NP-Hard problem. LEACH-C makes use of Simulated Annealing algorithm to address this problem. LEACH-C is more efficient than LEACH. LEACH-C delivers about 40% more data per unit energy than LEACH because the BS has global knowledge of the location and energy level of all nodes in the network. Also LEACH-C always insures that there will be always K optimal number of cluster heads in every set-up phase which is not ensured in LEACH.

Limitation of this algorithm is that CH selection and cluster formation will be done by the BS. So there is more load on the BS and BS must have information about the energy level of every nodes. After each round, energy information and location of every node should be known to the BS. Another limitation is that if sensor nodes are mobile nodes then nodes must be equipped with the global positing system or some location finding algorithms.

O. Younis et al [10], proposed a hierarchical, distributed, clustering scheme. In Hybrid Energy Efficient Distributed (HEED) inter-cluster communication is done in multi-hop manner and intra-cluster communication is done in single-hop manner. CH selection is done on the basis of two parameters: residual energy and intracluster communication cost. Residual energy is used to probabilistically choose the initial set of CHs and intracluster communication cost is used to determine the node degree or node’s distance with its neighbors. When clusters are formed node will select that CH which requires less intracluster communication cost. HEED mechanism provides uniform distribution of CHs across the network. The CH selection procedure takes constant number of iterations. Every sensor continues this procedure until it finds the CH with least cost. Before selecting as a permanently CH, the sensor will become tentative CH. The node will be a tentative CH if its probability is less than 1. And this node will become permanent CH if its probability value reaches to 1. After becoming CH it will send announcement message. If sensor node does not get any announcement message from any CH, the sensor elects itself to be a CH and then sends an announcement message to its neighbors informing them about the change of status. Each sensor doubles its probability value and goes to the next iteration. It stops executing this phase when its probability value is 1.

HEED can be applied to sensors which require scalability, prolonged network lifetime, fault tolerance and load balancing. HEED improve the energy consumption and network lifetime compare to LEACH.HEED support long rang communication and it provides uniform distribution of CHs. HEED has limitations of hot spot problem. In hot spot problem nodes that are near to the BS will die earlier because they have more data to transmit. Energy consumption is also not balanced between CHs. There are more overheads because cluster formation requires more iteration. This is complex algorithm.

S. Murguganathan et al [6], proposed a centralized routing protocol. In Base-station Controlled Dynamic Clustering Protocol (BCDCP) the base station will make decision about CHs, routing paths and rotation of CHs. The main focus was on creating balanced cluster that is cluster with equal number of members and uniform placement of CHs across the network. BCDCP operates in two phases setup and data communication phase. Setup phase includes cluster setup, CH selection, CH-to-CH routing path information and schedule creation for each cluster. During setup phase, all nodes send their information about current energy status to the BS. Based on these energy levels, BS will compute average energy level and create one set of nodes whose energy is more than this average energy. The CHs will be selected from this set only and it ensures that only nodes with sufficient energy will be selected as CH.

This algorithm solves the hot spot problem because the nodes that are transmitting data to the BS are changed in every round. So, there is not a single fix node whose energy will be used more and died earlier. CH selection overhead is reduced. Cluster splitting algorithm was used to form the clusters If balanced clustering technique is used to form the cluster then it will create the balanced cluster but it will not be energy efficient.

B. Tang et al [7], proposed a Geographic Energy Aware Routing Centralized Clustering (GEAR-CC) algorithm. Being a centralized clustering algorithm it utilizes the high power of BS to conduct every transmission in WSN. BS formulates the optimal transmitting schemes for all sensor nodes by using global information of topology and energy. The optimization is achieved by making trade-off between energy cost and node's residual power. Sensor nodes don't care about routing protocols. They just need to forward data to the next node with specified ID. Sensor nodes don't know to which cluster they belong. It ensures that the route used to collect data is optimal in global. Each node maintains an integer locally which refers to the next hop node's ID.

In GEAR-CC, each data collection task can be broken into three steps. They are best-route finding, next-hop setting and data transmission. The best-route finding operations are completed in BS. Any cluster head is selected randomly to transmit the data to the BS. So, it solves the hot-spot problem. All the routes are globally optimized because best route is found by BS for each node. But limitation of this algorithm is that inter-cluster communication cost is not optimized. Clusters are formed only on the basis of geographical position of nodes in the sensor network and energy parameter is not considered for CH selection.

D. Weigh et al [8], proposed a distributed clustering algorithm. Energy-Efficient Clustering (EC) algorithm determines the cluster size depending upon hop distance to BS. It achieves approximate equalization of node lifetimes and reduces energy consumption. Each sensor node makes observations and based on this a single data packet is produced and then transmits the packet to its associated CH. Each CH node collects the observation packets from its associated member nodes and a single summary packet is produced by combining data from cluster members. This summary packet represents the cluster. . This three-step process is referred to as a single data collection round (DCR) of the entire WSN operation.

To determine the CH candidates, a probability scale is assigned to each sensor. According to this value, each sensor decides on becoming a CH-candidate. Upon being selected as a CH-candidate, each CH-candidate transmits a CH-advertisement packet and advertises its residual energy level within a neighborhood. A CH-candidate monitors advertisements from other candidates and it will not act as a CH if a higher energy level is reported by another. Eventually, the candidates with the highest residual energy among their neighboring CH-candidates become the CHs during that particular DCR. It enables the choice of the actual CH nodes to be based on the most recent sensor energy stocks. After the CHs are elected, each CH transmits a CH-announcement packet within an area of transmission radius and informs other sensors of its availability as a CH. Each sensor may collect announcement packets from multiple CHs and selects the CH that has generated the announcement packet with the highest RSSI as the ideal CH to associate to. Nodes associate to CHs via sending a CH-association request and upon reception of a subsequent CH-confirmation.

This algorithm is energy efficient and solves the hot-spot problem. But CHs are selected on the basis of average energy not on the basis of residual energy.

R. Randriatsiferana et al [11], proposed algorithm to improve the performance of LEACH. This algorithm changes the criteria for electing a CH and the determination of optimal number of CHs. The optimal number of CHs is based on the density of the covering. Nodes having the highest remaining energy and the lowest energy variance consumption becomes cluster-heads. The variance parameter keeps energy consumption dispersion. This approach provides optimal number of clusters and it is energy-efficient.

The variance parameter keeps energy consumption dispersion, if the considered node is elected as cluster-head. This dispersion highly depends on the relative positioning of the node to the base station. This is useful for predicting node status when elected as cluster-head in the current round in order to balance the network energy.

M. Dhanaraj et al [12], proposed an algorithm named Optimal Placement of Cluster-heads (OPC) to solve problem of hot-spot. In order to balance the lifetime of the CHs, the CH placement algorithms should consider the fact that the node density and their transmission ranges are related parameters. This algorithm formulates an optimization problem and finds the optimum number of CHs to be deployed and their optimum transmission ranges for the given network parameters.

This algorithm provides the optimum density and radio transmission range for the cluster heads in order to achieve the maximum network lifetime with the minimum network cost. This algorithm also balances the lifetime of the cluster-heads so that the whole energy available at each cluster-head can be utilized. Additional cluster-heads are placed near the sink to handle the high load and hence, the density of the cluster-heads decreases as the distance increases between the sink and the cluster-head. As the cluster-head density increases, the transmission range of the cluster-heads can be reduced in order to save energy. Limitation of this algorithm is that receiving power and noise is not optimized.

III. COMPARISON OF EXISTING CLUSTERING ALGORITHMS

This section shows comparison of different schemes based on parameters as follows. Data transmission is used to determine whether data can be transmitted in single hop or multi hop. In single hop, data will be directly transmitted to the BS. While in multi hop, data will be transmitted to the intermediate node and then it will be transmitted to the BS. Energy consumption is used to determine how much energy is consumed by entire network. Load balancing parameter determines whether load of entire network is distributed between every node of the network or not. Approach refers if the algorithm is centralized or distributed. in centralized approach all the decisions regarding CH selection and cluster formation is done by the BS while in distributed node itself will decide whether to become CH or not and which cluster it should join.

TABLE 1 COMPARISION OF EXISTING CLUSTERINF ALGORITHMS

Algorithms	**Parameters**			
	Data Transmission	***Energy Consumption***	***Load Balancing***	***Approach***
LEACH[10]	Single Hop	More	Moderate	Distributed
LEACH-C[6]	Single Hop	More	Moderate	Centralized
HEED[11]	Multi Hop	Less	Moderate	Distributed
BCDCP[7]	Multi Hop	Less	Good	Centralized
EC[9]	Multi Hop	Less	Good	Distributed
GEAR-CC[8]	Multi Hop	Less	Moderate	Centralized
K-LEACH[12]	Multi Hop	Less	Good	Centralized
OPC [13]	Multi Hop	More	Good	Distributed

IV. CONCLUSION

Clustering protocol is most suitable for large scales WSN. Currently there are various clustering algorithms are available and only some of them is reviewed in this paper. This paper also presents the evaluation of the same on the basis of different parameters. Moreover, limitations and advantages of each algorithm is also discussed.

REFERENCES

[1] I. Akyildiz, Y. Sankarasubramaniam, E. Cayirci, “Wireless sensor networks: a survey”, Computer Networks, Vol. 38, Issue 4, PP. 393-422, 2002.

[2] J. Karaki, A. Kamal, “Routing techniques in wireless sensor networks: a survey”, IEEE Communications, Vol. 11, Issue 6, PP.1536-1284, 2002.

[3] Y.Jennifer, M. Biswanath, G.Dipak, “Wireless sensor network survey”, Computer Networks, Vol. 52, Issue 12, PP. 2292-2330, 2008.

[4] M. Lee, S. Lee, "Data Dissemination for Wireless Sensor Networks", in Proceedings of the 10th IEEE International Symposium on Object and Component Oriented Real-Time Distributed Computing, 2007.
[5] S. Prasanam, S. Rao, "An Overview of Wireless Sensor Networks Applications and Security", International Journal of Soft Computing and Engineering, Vol. 2, Issue 2, 2012.
[6] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An Application Specific Protocol Architecture for Wireless Microsensor Networks", Wireless Communication, IEEE, Vol. 1, Issue 4, PP. 660-670, 2002.
[7] S.Muruganathan, F. Daniel, R.Bhasin, A.Fapojuwo, "A Centralized Energy-Efficient Routing Protocol for Wireless Sensor Networks", Communications Magazine, IEEE, Vol. 43, Issue 3, PP. S8-13, 2005.
[8] B. Tang, D.Wang, H.Zhang, "A Centralized Clustering Geographic Energy Aware Routing for Wireless Sensor Networks", in proceedings of SMC' 13, IEEE International Conference on Systems, Man, and Cybernetics, Washington, DC, USA, PP. 1-6, 2013.
[9] D. Wei, Y. Jin, S. Vural, K. Moessner, R. Tafazolli, "An Energy-Efficient Clustering Solution for Wireless Sensor Networks", Wireless Communication, IEEE, Vol. 10, Issue 11, PP. 3973-3983 ,2011.
[10]W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Micro sensor Networks", Proceedings of the 33rd Hawaii International Conference on System Sciences, IEEE, PP. 4-7 Jan. 2000.
[11]O. Younis, F. Sonia, "HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks", Mobile Computing, IEEE, Vol. 3, Issue 4, PP. 366-379, 2004.
[12]S. Rivo, F. Alicapa, R. Lorion, A. Mohammed, "A Clustering Algorithm based on Energy Variance and Coverage Density in Centralized Hierarchical Wireless Sensor Networks", IEEE, Africon, 2013.
[13]M. Dhanaraj, C. Murthy, "On Achieving Maximum Network Lifetime Through Optimal Placement of Cluster-Heads in Wireless Sensor Networks", IEEE International Conference on Communication, PP. 3142-3147, Glasgow, 2007.