International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2020 IJSRCSEIT | Volume 6 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT2062142

Improvement in K-medoid Clustering using Density based Node Selection

Biresh Kumar[1], Pallab Banerjee[2], Amarnath Singh[3], Anurag Kumar[4], Avinash Kumar[4]

[1,2,3] Assistant Professor, Department of Computer Science and Engineering, Amity University Jharkhand, India

[4,5]B. Tech Scholar, Department of Computer Science and Engineering , Amity University Jharkhand, India

ABSTRACT

The wireless sensor networks have become a very hot topic of research in the recent years because of their wide range of applications such as industrial and agricultural monitoring, military surveillance, smart homes etc. These sensors can also be used in locations that are potentially hazardous to the human beings or are just out of our reach at this moment of time.

A wireless sensor network is collection of large number of individual sensor nodes. A sensor usually comprises of components for sensing the environment, processing, communicating and a power supply. The major limiting factor of the wireless sensor networks is the limited amount of power that each of the sensors can carry and also the energy in these sensors cannot be replenished easily. So, the major design challenge in a wireless sensor network has always been about reduction in the consumption of energy by the sensors. Clustering is one of the popular methods to reduce energy consumption in wireless sensor networks. Here, we propose a scheme to decrease the energy consumption and prolong the lifetime of wireless sensor network. The main idea behind the scheme is that we try to minimise the communication distance between the individual sensor nodes using the clustering technique.

Keywords : K-medoid Clustering, Low Energy Adaptive Clustering Hierarchy

I. INTRODUCTION

The wireless sensor networks usually comprise of hundreds and thousands of individual sensors equipped with computing, sensing and communicational abilities. Each node can sense a part of the environment, perform easy computational tasks on the sensed data and communicate this data with the peer nodes or directly to the external base station. The base station can be considered similar to a sensor with increased amounts of computational power, memory and power supply. The deployment of the sensors can be carried out either in random manner or manually to achieve better positioning of individual sensors.

These wireless sensor networks provide a cost effective, low maintenance and persistent way of collecting data from a variety of different sources. The sensors are equipped with a very limited power supply and they need to operate for a long period of time in order to achieve its objectives. So different techniques must be used to eliminate the energy inefficiencies that shorten the lifetime of the network. Some real-life applications of these networks include weather monitoring, industrial and agricultural monitoring, intrusion detection, field imaging, security surveillance etc. In most of the scenarios the sensors are deployed in random manner dispersed in a large area. In such cases, some regions covered by the network are more densely populated by sensors than other regions thus increasing the chances that multiple sensors in network are sensing the same data. The energy used in acquiring this redundant data and transmitting is wasted.

In this paper we propose a new K-Medoids based clustering scheme to increase the lifetime of the network. Every node calculates its local sensor density in order to determine whether it will take part in the clustering process or not.

II. RELATED WORK

WSNs have severe energy constraints where the network needs to operate for a long period of time unattended. So, in order to optimize the energy consumption and improve the energy utilization rate in the sensor nodes, a number of researches have been carried out. These can be broadly classified into two categories: tree-based routing algorithms and cluster-based algorithms. In the tree-based methods, the formation of an optimal routing tree is a really hard problem to solve. Cluster-based methods provide better node management but the cost of the maintenance of the clusters is quite large.

Low Energy Adaptive Clustering Hierarchy (LEACH) proposed by Heinzelman is a classical cluster-based algorithm. In LEACH, the network is divided into several clusters. A cluster head node is selected for each cluster by probability. The rest of the common sensor nodes form the cluster by choosing the nearest cluster head node. The common sensor nodes sense the environment around them and transmit this data to the cluster head node. The cluster head node aggregates the data collected from all sensor nodes and transmit the compressed data to Base Station. LEACH reduces the energy consumption WSNs but the random probabilistic selection of head node may result in a poor cluster formation. The uneven distribution will disrupt the load balance in the network and decrease its efficiency.

LEACH-centralized was proposed to overcome the inefficiencies of LEACH algorithm by applying a centralised clustering algorithm. In LEACH-C, the BS receives the data about the position and residual energy from all the sensor nodes. Using this information, the BS calculates the number of clusters as well as the set-up network.

1. System Model

A. Network Model

The system comprises of a large number of sensor nodes and one base station (BS). All sensors can be categorized in two ways. One is common node and the other is head node. These nodes are dispersed in a region having area A m^2. The task of the common nodes is to sense the environment and transmit the received data to the cluster head node. The cluster head node is selected from the common nodes and is responsible for aggregation of the sensor data and transmission to BS.

B. Energy Model

The general radio model is used as the energy model. The energy consumed during the communication phase is the most taxing, hence we only consider the energy consumed during the communication phase. The total energy consumption consists of energy expended during the data transmission and receiving phase. The energy consumed during the exchange of L-bit data can be calculated by equation 1.

$$E_{Tx}(L,d) = E_{Tx}(L) + E_{mp}(L,d) = \begin{cases} LE_{Tx} + ke_{fs}d^2, & d < d_0 \\ LE_{Tx} + ke_{mp}d^4, & d \geq d_0 \end{cases} \quad (1)$$

$$E_{Rx}(L) = LE_{Rx} \quad (2)$$

where E_{Tx} and E_{Rx} are the per bit energy dissipations for transmission and reception, respectively, e_{fs} *and* e_{mp} denote the transmit amplifier parameters corresponding to free space and multipath propagation models respectively and d_0 is the threshold distance given by

$$d_0 = \sqrt{e_{fs}/e_{mp}} \quad (3)$$

2. Our Proposed Scheme

In order to reduce the energy consumption and to extend its lifetime, we proposed a K-Medoids clustering scheme which considers the local density of the sensors in order whether or not the sensor will participate in the cluster formation. In the proposed scheme, each cluster is formed using the K-Medoids scheme i.e. there is a cluster head (CH) and member nodes whose cluster membership depends on the local sensor density. The chances of a certain selection of nodes joining a cluster is dynamically adjusted based on the local node density. By the prevention of transmission of redundant data, the network lifetime is generally increased. In this scheme, the algorithm consists of three phases: pre-setup phase, set-up phase and the communication phase.

A. Pre-Setup Phase

To compute the local density of a sensor node, we first calculate the average area acquired by each cluster which is gives as follows

$$A : N = \pi R^2 : \frac{N}{k} \quad (4)$$

Where A is the total area of the network, N is the total number of sensor nodes over the network, R is the transmission radius of a cluster and k is the number of clusters. Using equation (4), we calculate the radius of a cluster given as

$$R = \sqrt{\frac{A}{\pi k}} \quad (5)$$

Every node i in the network calculates D_i, the local density of the i^{th} node. D_i indicates the proximity of the other nodes that exists around node i. For the calculation of D_i, each node broadcasts a discovery packet to all other nodes in its transmission radius. By counting the discovery packets from adjacent nodes, each node calculates the local density of nodes around itself.

If the local density around a certain node is higher than the average density, it indicates that there are a large number of nodes that measure the same value. The recommended number of nodes per cluster or the average number of nodes in a clutter is given as

$$D = \left\lceil \frac{N}{k} \right\rceil \quad (6)$$

Where N is the total number of nodes and k is the number of clusters formed.

If the local density around any node i is less than D i.e. $D_i < D$, it indicates that the node i always joins the cluster for current round. Otherwise the node i joins the cluster with a probability of D/D_i. The probability that the node i joins the cluster is given as

$$P(i) = \begin{cases} \mu \times \frac{D}{D_i}, & D_i > D \\ 1, & otherwise \end{cases} \quad (7)$$

Where μ is the adjusting factor. If a certain node in a high-density region takes part in the clustering making process in a round. Then the probability of that node for the next round of cluster formation decreases.

B. Setup Phase

The set-up phase is associated with partitioning of the network into appropriate clusters. The clustering process is carried out using the K-Medoids scheme.

The number of clusters in the network is given as

$$k = \sqrt{\frac{N}{2}} \quad (8)$$

Where N is the total number of sensors and k is the number of clusters.

Now k random points are selected as the initial medoids. Each node is associated with the closest medoid. In the next pass, a new node is selected as medoid, if the cost to connecting to this node more than the previous medoid then the previous medoid is swapped with the new one. The same process is repeated until we get recurring medoids. The cost in K-Medoids scheme is given as

$$C = \sum_{Ci} \sum_{Pi \in Ci} |Pi - Ci| \quad (9)$$

Unlike the K-Means algorithm, K-Medoids uses real node to represent a cluster head rather than a theoretical mean position.

In the next iteration of clustering, a greedy approach is applied for selecting the new cluster head to replace the current cluster head node. The average residual energy of the network is calculated as

$$E_{av} = \frac{\sum_{i=1}^{N} E_{res}}{N} \quad (10)$$

Where E_{res} is the residual of each node in the network and N is the total number of sensors.

Let $I = \{I_1, I_2, \ldots \ldots \ldots \ldots \ldots, I_k\}$ be the set of nodes. We select a common node I_{random} to replace the cluster head node. The residual energy of this random node must be higher than average residual energy of all nodes i.e. $E_{I_{random}} > E_{av}$ and I_{random} should have the minimum inter-nodal distance of all qualifying nodes. The inter-nodal distance is given as

$$S = \sum_{x \in C_i} dist(x, I_{random}) \quad (11)$$

Where S is the internodal distance

C. Communication Phase

After the formation of the required clusters, the cluster head node prepares a TDMA schedule and each node is allotted a time slot. This schedule is broadcasted to all the nodes in cluster. During the communication phase, the sensor nodes begin sensing and transmitting data to cluster heads.

The cluster head node receives all the data, aggregates it and sends it to the base station. After a predetermined amount of time, the network goes back into the setup phase again and enters another round to select new cluster heads.

III. CONCLUSION

In this paper, we describe a clustering protocol for Wireless Sensor Networks which minimizes energy consumption by dynamically adjusting the number of nodes that take part in the cluster formation based on local density of node. In regions which have a high density of sensors, multiple sensors sense and transmit the same data. The energy consumption is hence reduced by preventing the transmission of this redundant data.

IV. REFERENCES

[1]. Banerjee, S., Khuller, S.: A Clustering Scheme for Hierarchical Control in Multi-hop Wireless Networks.

[2]. In: Proceedings of IEEE INFOCOM. IEEE Press, Los Alamitos (2001).

[3]. Kwon, T.J., Gerla, M.: Clustering with Power Control. In: Proceeding of MilCOM 1999 (1999).

[4]. Bandyopadhyay, S., Coyle, E.: An Energy-Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks. In: Proceedings of IEEE INFOCOM (2003).

[5]. J.Y. Chang and P.H. Ju, An efficient cluster-based power saving scheme for wireless sensor networks", EURASIP Journal on Wireless Communications and Networking, May 2012. J.Y. Chang and P.H. Ju, An efficient cluster-based power saving scheme for wireless sensor networks", EURASIP Journal on Wireless Communications and Networking, May 2012.

[6]. W. Wang, S. Zhang, G. Duan, et al. Security in wireless sensor networks, IEEE Wireless Communications, 2008.

[7]. Amis, A.D., Prakash, R., Vuong, T.H.P., Huynh, D.T.: Max-Min D-Cluster Formation in Wireless Ad Hoc Networks. In: Proceedings of IEEE INFOCOM (2000).

[8]. O Younis, S Fahmy, HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE Trans Mob Comput.

Cite this article as :

Biresh Kumar, Pallab Banerjee, Amarnath Singh, Anurag Kumar, Avinash Kumar, "Improvement in K-medoid Clustering using Density based Node Selection", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 2, pp. 490-494, March-April 2020. Available at doi : https://doi.org/10.32628/CSEIT2062142
Journal URL : http://ijsrcseit.com/CSEIT2062142