

A New Approach for Big Data Gathering in Dynamic Wireless Sensor Networks

Antu Raj S¹, Sangeetha Jose²

¹Department of Information Technology, Government Engineering College, Idukki, India

²Department of Information Technology, Government Engineering College, Idukki, India

Abstract: Big data refers to huge amount of data that we can't handle using our traditional programming concerns. Wireless Sensor Networks is one of the main contributors of big data. Data dissemination and data gathering are two important operations in wireless sensor networks. Mobile sink approach is an efficient way to gather data in wireless sensor networks. Earlier works [1][6] have considered the mobile sink approach for energy efficient data gathering. In that works, Mobile sinks gather data from centroid of each cluster in the wireless sensor networks. However this approach considered distance as the parameter for clustering and centroid calculation which is not possible for a dynamic wireless sensor networks. Hence a new approach is needed for the clustering and cluster head selection in dynamic wireless sensor networks which in turn helps for the big data gathering. Modified clustering algorithm with mobile sink based data gathering is proposed in this paper for clustering and big data gathering in dynamic wireless sensor networks. The proposed algorithm can gather data from dynamic wireless sensor networks in an energy efficient manner.

Keywords: Big Data, Wireless Sensor Networks, Mobile Sink, Static Sink

1. Introduction

Big data represents the data that has high velocity, variety and volume of data which is difficult to gather using the currently available technologies. As per IBM's definition Big data is a combination of five 'V' that is five properties. Velocity, Veracity, Volume, Variety, Value as shown in Figure.1

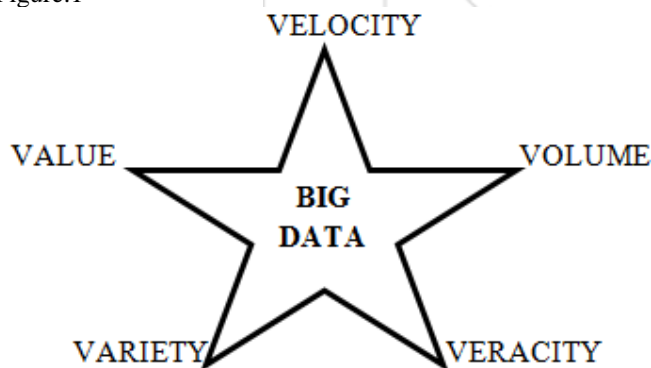


Figure 1: Properties of Big Data

Major contributors of big data are wireless sensor network. Wireless sensor network (WSN) is nothing but a collection of sensor nodes that are arranged in a distributed manner to sense the environment in which it deploy on. The sensed data are collected for monitoring purposes. Usually sensors collect information like temperature, humidity, pressure and further give this data to a sink. Wireless sensor networks are usually deployed for monitoring the environments and act accordingly. Wireless sensor network consist of mainly three components sensor nodes, sink node and base station. Sensor node senses and collects values/information from its surroundings. The sink node collects these values from the sensor node and gives it to the base station.

When Wireless sensor network is used for data gathering, the biggest constraint is the energy consumption. Because of the

low power nature of sensor nodes and energy consumption nature of wireless transmission, an energy efficient approach is needed to gather the big data from wireless sensor network. In order to overcome this challenge, Mobile sink approach is used for data gathering from sensors. In this approach, Mobile sinks are used instead of static sinks for collecting data from the sensor nodes. As a result when the sink node moves towards the sensing area, the sensor nodes send data to the sink node when the sink node comes in their transmission range. Hence the energy consumption can be reduced to a significant level.

1.1 Motivation

Wireless sensor networks can gather data in an energy efficient manner if it uses the mobile sink approach. The process of data collection will become easier by this approach. The previous works [1][6] also considered mobile sink approach to gather data in an energy efficient manner. For data gathering, the network is divided into clusters and mobile sink will gather data from centroid of each cluster. The centroid is nothing but the location at which mass of a cluster is concentrated. It proposed EM (Expectation Maximization) algorithm for clustering in which clustering and cluster centroid selection depends on the wireless communication distance. Even though it addressed the data gathering issue in wireless sensor networks, it does not focus on the mobility of sensor nodes. However, for a dynamic sensor network, the distance parameter can't be considered for clustering since the value of distance constantly changing. Since Wireless sensor networks are collection of sensor nodes which are communicated in wireless manner, the term dynamics has an important role. Therefore, it is equally important to focus on the mobile sink approach in the dynamic sensor environment. Hence we proposed a new algorithm for gathering data from the dynamic sensor networks in an energy efficient manner. This algorithm can

play an important role in both big data gathering and wireless sensor network

use of network model

1.2 Organization of Paper

This paper is organized as follows. Section 2 gives the literature review which describes various clustering algorithms in wireless sensor networks. Section 3 gives the system design in which existing and proposed systems are analyzed and proposed algorithm is explained. Section 4 gives the conclusion of the proposed system and future works.

2. Related Works

Wireless sensor network is network of sensor node which is distributed in wireless manner. Wireless sensor network consist of sensor nodes and base station. These types of networks are deployed for monitoring an environment or system. Limited energy and bandwidth are two constraints of wireless sensor networks. An energy-efficient, scalable approach to reducing the energy consumption over the whole network is the clustering process. It also increases network lifetime. When clustering is applied in wireless sensor network, nodes are divided into multiple clusters, and a head node will be assigned for each cluster as a leader, which collects all the data from the sensors belonging to its cluster and relays these data to the sink node directly or by multiple hops[6]. Different methods and algorithms are proposed for clustering in wireless sensor networks. Some works are reviewed and summarized here.

In [2] Dynamic Clustering Routing Considering Load (DCRCL) is proposed for clustering. It addresses the problem associated with leach. In leach, there is a chance to select a low energy node as cluster head. Another problem is that communication between cluster head and base station is in a single hop manner which will consume more energy and thus reduce the network life cycle. Even though large numbers of algorithms like PEGASIS, HEED are developed to overcome this problem, data delay is high and it is not suitable for large sized networks. Even though these algorithms can improve the total lifetime of the network, the load problem is not addressed. In DCRCL the load problem is addressed. The algorithm consists of mainly three phases. Cluster head selection, Cluster set up and inters cluster routing. Residual energy and node load is considered for cluster head selection. In cluster set up phase in order to avoid the unbalanced distribution of cluster heads in LEACH, cluster head competitive mechanism is used. That is after being candidate cluster head, the node competes with other candidates in its communication range and the one with highest residual energy will be selected as cluster head. In the third phase, inter cluster routing it addresses the problem with LEACH. In Leach data transmitted directly from cluster head to base station which reduce the life cycle of cluster head as well as network. In DCRCL, energy balanced clustering communication model is established for efficient

In [3] Prasanth Krishnan et al. compared the performance of different static as well as dynamic clustering protocols. The performances are evaluated using various node distribution schemes like Random node distribution, Poisson distribution and Uniform distribution. Routing is the process of determining the path between source and destination. There are many types of routing protocols on the basis of different parameters. When it comes to wireless sensor networks, the routing protocols are mainly classified into flat, hierarchical based routing and location based routing. In hierarchical, nodes in the network have different roles. It can be further classified in to two, static and dynamic. In dynamic, cluster are dynamic in nature but in Static, once the clusters are created remain same throughout network lifetime. Prasanth et al. compare the performance of LEACH, Enhanced Energy-Efficient Protocol with Static Clustering (E3PSC) and Energy-Efficient Protocol with Static Clustering (EEPSC). The Low Energy Adaptive Clustering Hierarchy (LEACH) is Dynamic Hierarchical Routing for sensor network. In LEACH the selection of cluster heads will be in a random manner. Cluster head will compress and aggregate data that are coming from cluster nodes and send this data to the base station. LEACH uses a TDMA/code-division multiple access (CDMA) MAC to reduce inter-cluster and intra-cluster collisions. When constant monitoring to the network is needed LEACH is the most appropriate solution. In LEACH there are two phases, the setup phase and the steady phase. In the set up phase, the cluster organization and cluster head selection takes place. In the steady state phase, the actual data is transferred to the BS. For minimizing the overhead the setup phase is less durable as compared to steady phase. The problem with LEACH is that it is not applicable to large networks. It also assumes that nodes always have data to send, and nodes located close to each other have correlated data. Performance analysis depicts that nodes over Poisson random distribution are alive for maximum network lifetime. Network lifetime of Nodes deployed on uniform random number distribution is less than network lifetime of nodes deployed on Poisson distribution. The second protocol, Enhanced Energy-Efficient Protocol with Static Clustering (E3PSC) is an improvement of an existing routing scheme, Energy-Efficient Protocol with Static Clustering (EEPSC). In this protocol, Spatial distribution of sensors nodes in network and their residual energy are considered for selecting cluster head. This will reduce the communication overhead within cluster. Performance analysis shows that E3PSC will provide good result in uniform random distribution. In Energy-Efficient Protocol with Static Clustering, it eliminates the overhead of dynamic clustering by dividing the network into static clusters. It distributes the energy load among high power sensor nodes by using temporary-cluster-heads. In effect it extends network lifetime. There are three phases in EEPSC, set-up phase, responsible node selection phase and steady-

Name of Paper	Description	Merits	Demerits
Research on Dynamic Clustering Routing Considering Node Load	Proposed an clustering algorithm that con-siders network load and residual	* Efficient * Addresses the problems of LEACH	* Nodes are static

for Wireless Sensor Networks[Yi Sun, Can Cui, Shanshan Ke, Jun Lu, 2013]	energy	* Increases the life cycle of network * Increases the data volume of network	
Comparison and Performance Analysis of Dynamic and Static Clustering Based Routing Scheme in Wireless Sensor Network[Prasanth Krishnan, 2013]	Compared the performance of different static as well as dynamic clustering protocols.	* Using various node distribution schemes used for evaluation	
Weighted Dynamic Distributed Clustering Protocol For Heterogeneous Wireless Sensor Network[Said Benkirane, 2012]	Proposed an autonomous, efficient and dynamic algorithm for clustering	*High Stable time * High efficiency * Less energy consumption * Increased network lifetime * Heterogeneous WSN	* Sink is static * Sensor nodes are static
Dynamic Clustering Protocol for Data Forwarding in Wireless Sensor Networks[deepali virmani]	Proposed a dynamic cluster formation method	* Increased network life time * Enhanced energy conservation	* Delay * Didn't support real time parameters

state phase. Static clustering Scheme and idea of temporary cluster heads are the advantages of this protocol.

In [4] Benkirane et al. proposed WDDC [Weighted Dynamic Distributed Clustering], which is based on the ratio of residual energy and average energy of the network. This approach takes into consideration distances between nodes and the base station to determine near nodes and distant nodes in order to give more chance to the nearest nodes to be cluster heads by modifying the election probability value for every type of nodes. WDDC is dynamic approach and it is autonomous as well as more energy efficient. It increases the lifetime of networks. In WDDC, nodes with more energy than the other nodes and the nodes with less distance from the BS have more chance to be selected as a cluster-head for current round. So this approach introduced a concept of weighted probabilities for every type of nodes according to their residual energy and the average energy of the network in current round. The distance between nodes and the base station are also taken into consideration for favoring the nodes with more energy and closest to the BS to become cluster heads. The method of communication between cluster heads and cluster nodes are similar to LEACH. That is After the cluster head formation, it will inform all other nodes about its selection as cluster head. After that other nodes choose the most appropriate cluster heads and organize themselves into local clusters. Then cluster head gets data from the cluster nodes according to TDMA Schedule. When it comes to the communication between cluster head and base station, after receiving all the data from the cluster nodes, the cluster head node executes signal processing functions to compress the data into a single signal. Then each cluster head send the combined data to the base station. The consumed energy of cluster head is composed of three parts: data receiving, data aggregation and data transmission. The main advantage of WDDC is that it have significantly high stable time which increases the efficiency of networks. In addition

WDDC consumes less energy when compared to other protocols. So network lifetime can be increased.

Deepali virmani et al. proposed a dynamic cluster formation method where clusters are refreshed periodically based on residual energy, distance and cost in [5]. Here distance is considered in order to minimize energy consumption and increase lifetime of the network. The main focus is to minimizing energy consumption, maximizing network lifetime, and real time communication. A dynamic clustering protocol (DSP) is proposed which forms the cluster based on the residual energy of each node. The concept of assigning different energy levels to different nodes is introduced for balancing the responsibility among the cluster nodes. The node with the highest energy level looks for nodes within its transmission range forms a cluster and appoints itself as the cluster head of the cluster formed. Then data transmitted from all the cluster nodes to the cluster head. That is cluster head act as a data aggregating node in for its respective cluster. After sending data to the cluster head nodes will go to sleep state until they have something more to transmit. Only Cluster head communicate with other cluster head thus reduces the energy consumption. So in effect the lifetime of node increased. After data aggregation cluster head forward the data to base station. This step reduces the energy of the cluster head. So calculation of remaining energy and election of cluster heads are conducted in periodic intervals. Here clustering can be performed even though the position and energy of nodes are changing. This dynamic clustering will results in the effective energy utilization of all the nodes in the network. In short here proposed protocol is better in terms of energy conservation and enhancement of network life time as sleep and wait technology enhances the lifetime of nodes.

In [6], performance evaluation of wireless sensor networks for data gathering was done. This work analyzed the performance of wireless sensor networks with different number of mobile sinks. The clustering and centroid calculation was entirely based on the Expectation

Maximization algorithm. This paper considered the number of mobile sinks as an important factor which affects the performance of wireless sensor networks while performing data gathering. An equation was also derived which shows the relationship between number of mobile sinks required for a network with given number of sensor nodes.

3. Proposed System

3.1 System Design

The system is mainly designed to overcome the energy consumption issues during data gathering in Wireless sensor network. Mobile sink approach is used to solve this problem. The whole mass of a cluster is concentrated in centroid. The Sensors in wireless sensor network are organized as clusters. Mobile sink gather data from the centroid of clusters in WSN. Earlier work proved the fact that by using modified EM (Expectation Maximization) algorithm with mobile sink based clustering and data gathering will solve these issues and introduce a new relationship that will further reduce the energy consumption and cost effective with better performance. The expectation maximization method for clustering aims to minimize the sum of square of wireless communication distance. The main aim of this clustering scheme is to maximization of the expecting gain. For that it will use the distance property so it will reduce energy consumption by reducing the communication distance. That means nodes which are closest will form a cluster. Thus there are so many clusters are created in which nodes within cluster are of shortest distance than from the nodes between the cluster. In that case the sensor nodes were static, so distance calculation and application of EM algorithm was possible to do. But when it comes to our case, the sensor nodes are dynamic in nature. The previous approach is not applicable in the dynamic environment. The distance is not a static value in this case. So the EM algorithm can't be used here for clustering and centroid calculation.

Considered network topology is a heterogeneous one as shown in figure.2. Topology consists of mobile sink, base station and sensor nodes. Here the Sensor nodes are not homogeneous. They are heterogeneous in nature. That is some of them have high battery power while some not. There are two types of sink in the proposed topology, static sink and mobile sink. Static sink is nothing but the cluster head of each cluster. The basic idea is that Static sink will gather data from all of its cluster nodes and store it in the buffer memory, then the mobile sink will move towards the cluster head that is static sink of each cluster and gather /collect data from it. Later on the mobile sink will send this data to the base station.

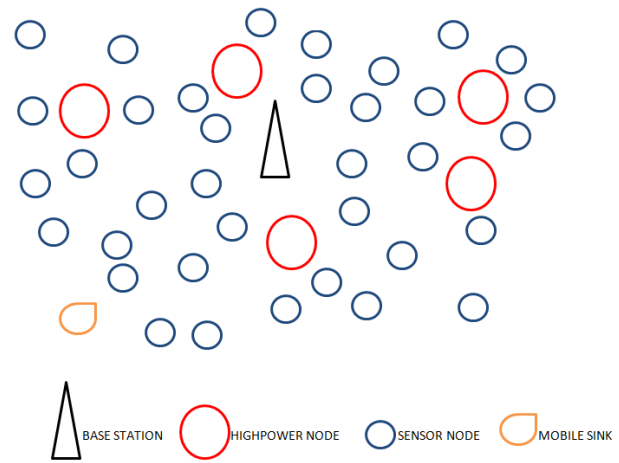


Figure 2: Network Topology

For energy efficient data gathering in dynamic wireless sensor network, here uses the two sink approach. The Figure 2. Network Topology clustering and cluster head selection is somewhat a difficult task since sensor nodes are highly dynamic in nature. The high power nodes are selected as cluster heads and works as a static node by gathering data from the sensor nodes. After the deployment of network, the low power nodes will broadcast their presence to all other nodes which are inside their proximity. When a high power node receives this broadcast, it will send a reply to the node which broadcasted the request. The node compares all replies from the high power nodes and chooses the highest power node as its cluster head. If it get reply from one node only, that high power node is selected as cluster head. Then the node notifies the cluster head about the selection. When cluster head node receives this notification about the selection, it will form a cluster with nodes from which it gets the cluster head notification. Cluster head of each cluster acts as static node. It gathers data from all the cluster nodes. The cluster head will count the no of packets it gather and when it reaches a limit it will inform the mobile sink to collect data from it. When mobile sink receives this notification it will move towards the cluster head and gather the data from it and then send it to base station. When a node leaves a cluster while moving it will send a notification to the cluster head. When all the cluster nodes has left from a cluster, cluster head will goes to a sleep state. The algorithm 1 explains the proposed algorithm for the clustering and cluster head /static sink selection.

Algorithm 1 Modified clustering algorithm with mobile sink based data gathering

- 1: Start
- 2: Deploy nodes in a heterogeneous manner
- 3: Each node broadcast its presence to all other nodes within its sensing area
- 4: When a high power node receives this broadcast, it will send a reply to the node
- 5: The node compares all replies from the high power nodes and choose the highest power node as its cluster head.
- 6: Node notify/inform the cluster head about the selection
- 7: When cluster head node receives this notification, it will form a cluster with nodes from which it gets the cluster head notification

- 8: Cluster head of each cluster acts as static Sink. It gathers data from all the cluster nodes.
- 9: The cluster head will count the number of packets it gather and when it reaches a limit it will inform the mobile sink
- 10: When mobile sink receives this notification it will move towards the cluster head and gather the data from it
- 11: Mobile sink sends data to base station
- 12: When a node leaves a cluster while moving it will send a notification to the cluster head and repeat 3-11
- 13: When all the cluster nodes has left from a cluster, cluster head will goes to a sleep state.
- 14: The mobile sink gather information until no more cluster head is to send data
- 15: Stop

3.2 System Implementation

The considered network consists of heterogeneous sensor nodes, mobile sink and base station. High power nodes will form cluster as explained in algorithm1 and act as static sink. The sensor nodes gather data from its surroundings. The static sink that is the cluster head of each cluster will gather data from sensor nodes and store it in the buffer memory. When the buffer memory is full, it will inform mobile sink. MS will visit the static sink and gather data from its buffer and send this data to the base station. The architectural framework for the implementation is shown in Figure.3

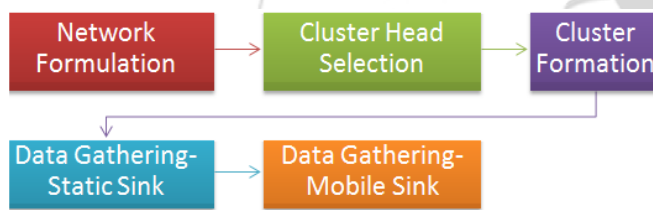


Figure 3: Network Architecture

Network formation phase is used to deploy the sensor node in the field where monitoring is to be performed. The network topology consists of sensor node, mobile sink and base station. Nodes are heterogeneous in nature. Some nodes have high battery power while others are normal nodes. In cluster head selection phase each low power node broadcast its presence to neighbors. The highest power node within the proximity of each node is selected as cluster head. The cluster formation phase is used for the formation of cluster. The cluster is formed by the cluster head i.e., cluster head form a cluster by with nodes from which it gets the cluster head notification. Then nodes within the cluster communicate with static sink when it senses data. When the buffer of static node is full it will notify mobile sink. Mobile sink then visit the static node and gather data from it. Hence data can be gathered from the dynamic wireless sensor network by using the proposed algorithm. Previous works [1],[6] considered static sensor environment for data gathering. The clustering and centroid calculation was performed using the Expectation-Maximization algorithm. However this paper proposed a new algorithm for big data gathering in dynamic wireless sensor networks in an energy efficient manner.

4. Conclusion

Examined the energy efficient data gathering method by using mobile sink approach in wireless sensor network. The existing approaches are not applicable to wireless sensor networks when sensor nodes are dynamic. Hence for gather data from the dynamic wireless sensor network two sink approach was proposed. Moreover, Modified clustering algorithm was proposed for the data gathering from dynamic wireless sensor networks. This mobile sink approach can also applied to homogeneous dynamic wireless sensor network which is considered as an open problem.

References

- [1] Daisuke Takaishi, Hiroki Nishiyama, Nei Kato, and Ryu Miura, "Towards Energy Efficient Big Data Gathering in Densely Distributed Sensor Networks", IEEE Transactions on Emerging Topics in Cloud Computing, February 2014
- [2] Yi Sun, Can Cui, Shanshan Ke, Jun Lu, "Research on Dynamic Clustering Routing Considering Node Load for Wireless Sensor Networks", Communications and Network, 2013, 5, 508-511
- [3] Prashant Krishnan, "Comparison and Performance Analysis of Dynamic and Static Clustering Based Routing Scheme in Wireless Sensor Network", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 4, April 2013
- [4] Said Benkirane, Abderrahim Beni hssane, Moulay Lahcen Hasnaoui, Mostafa Saadi and Mohamed Laghdhir, "Weighted Dynamic Distributed Clustering Protocol For Heterogeneous Wireless Sensor Network", International Journal of Wireless and Mobile Networks(IJWMN) Vol. 4, No. 6, December 2012
- [5] Deepali Virmani, Akshay Jain, Ankit Khandelwal, Divik Gupta, Nitin Garg, "Dynamic Clustering Protocol for Data Forwarding in Wireless Sensor Networks", International Journal of Computers and Technology, 2013
- [6] K R Remesh Babu, Suja G J, Philip Samuel, Sangeetha Jose "Performance Analysis Of Big Data Gathering in Wireless Sensor Network Using An EM Based Clustering Scheme", International Conference on Advances in Computing and Communications, September 2015
- [7] Tal Anker, Danny Bickson, Danny Dolev and Bracha Hod, "Efficient Clustering for Improving Network Performance in Wireless Sensor Networks", EWSN'08 Proceedings of the 5th European conference on Wireless sensor networks, 2008
- [8] P. Sasikumar, Sibaram Khara "K-Means Clustering In Wireless Sensor Networks", International Conference on Computational Intelligence and Communication Networks, 2012
- [9] A.B.M.Alim Al Islam, Chowdhury Sayeed Hyder, Humayun Kabir, Mahmuda Naznin, "Stable Sensor Network (SSN): A Dynamic Clustering Technique for Maximizing Stability in Wireless Sensor Networks", Wireless Sensor Network, 2010, 2, 538-554 .

- [10] Rashmi KR, Shivakumar AB, Ananda Babu J, “ Energy Efficient Clustering Techniques For Mobile Data Gathering In Distributed WSN ”, International Journal of Advanced Technology in Engineering and Science Volume No 03, Special Issue No. 01, March 2015

