A Project Report on

DESIGN AN ALGORITHM FOR WIRELESS SENSOR NETWORK

Submitted to the Department of Information Technology

For the partial fulfilment of the degree of B.E. in Information Technology

by

CHANDRIMA BHATTACHARYA (111308009)
SANKET CHAKROBORTY (111308056)
AKSHITA (111308058)
B.E., 4th year

Under the supervision of

DR. CHANDAN GIRI

Department of Information Technology

INDIAN INSTITUTE OF ENGINEERING SCIENCE AND TECHNOLOGY, SHIBPUR

May, 2017

Department of Information Technology
Indian Institute of Engineering Science and Technology, Shibpur

CERTIFICATE

This is to certify that the work presented in this report entitled "Design an algorithm for Wireless Sensor Network", submitted by Chandrima Bhattacharya, Sanket Chakroborty and Akshita having the examination roll number 111308009, 111308056 and 111308058 has been carried out under my supervision for the partial fulfilment of the degree of Bachelor of Technology in Information Technology during the session 2016-17 in the Department of Information Technology, Indian Institute of Engineering Science and Technology, Shibpur.

——————————————————

Dr. Chandan Giri
Assistant/Associate Professor
Department of Information Technology
Indian Institute of Engineering Science
and Technology, Shibpur

——————————————————

Dr. Arindam Biswas
Head of the Department
Department of Information Technology
Indian Institute of Engineering Science
and Technology, Shibpur

Date: 08/05/2017

Acknowledgements

We are thankful to Dr. Chandan Giri, Assistant professor, IIEST,Shibpur to help in making the project “Design an Algorithm for Wireless Sensor Network”. This project couldn’t had been completed without the able support of our mentor and the coordination of the members of the group.

Date: 08/05/2017

CHANDRIMA BHATTACHARYA
AKSHITA
SANKET CHAKRABORTY
Department of Information Technology
Indian Institute of Engineering Science
and Technology, Shibpur

ABSTRACT

A wireless sensor network is a collection of nodes organized into a cooperative network. Each node consists of processing capability (one or more microcontrollers, CPUs or DSP chips), may contain multiple types of memory (program, data and flash memories), have a RF transceiver (usually with a single omnidirectional antenna), have a power source (e.g., batteries and solar cells), and accommodate various sensors and actuators. The nodes communicate wirelessly and often self-organize after being deployed in an ad hoc fashion. Systems of 1000s or even 10,000 nodes are anticipated. Such systems can revolutionize the way we live and work.

Contents

1 INTRODUCTION

A wireless sensor network is an adhoc arrangement of multifunctional sensor nodes in a sensor field, disseminated to gather information regarding some phenomenon. Sensor nodes can be densely distributed over a large and may be remote area and collaborate their efforts to the benefit of the network to the extent that even if a number of nodes malfunction, the network will continue to function. There are two main layouts for wireless sensor networks. The first is a star layout where the nodes communicate, in a single hop, directly to the sink whenever possible and peer-to-peer communication is minimal. In the second, information is routed back to the sink via data passing between nodes. This multi-hop communication is expected to consume less power than single-hop communication because nodes in the sensor field are densely distributed and are relatively close to each other.

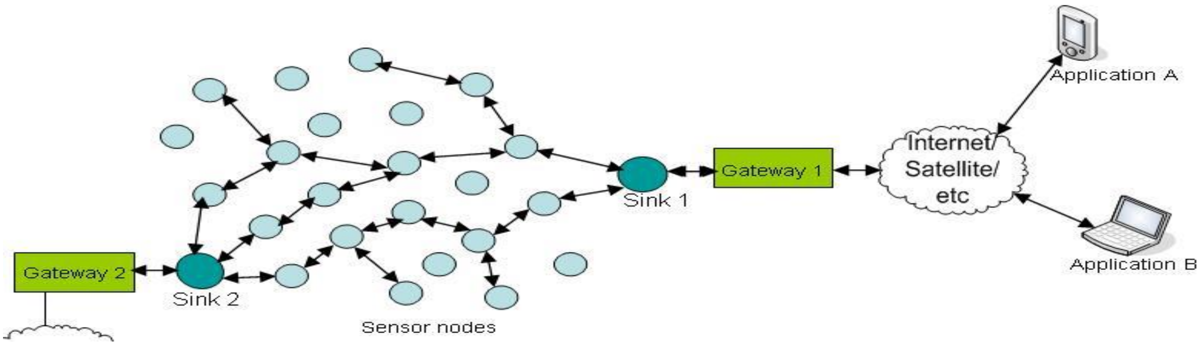

Fig 1.1 Diagrammatic representation of Wireless Sensor Network

Wireless Sensor Networks (WSN) integrate a large number of tiny devices with communication and processing capabilities aimed at several different applications, many related to monitoring (e.g. environmental, habitat, health or industrial monitoring). Typically in WSNs one of the main challenges is power limitation as the replacement of batteries is either costly or very difficult (for example when nodes are deployed in harsh environments).

2 PROTOCOL STACK OF WSN

A protocol stack for WSNs must support their typical features and singularities. The sensor network protocol stack is much like the traditional protocol stack, with the following layers: application, transport, network, data link, and physical. The physical layer is responsible for frequency selection, carrier frequency generation, signal detection, modulation and data encryption. The data link layer is responsible for the multiplexing of data streams, data frame detection, medium access and error control. It ensures reliable point-to-point and point-to-multipoint connections in a communication network. The network layer takes care of routing the data supplied by the transport layer. The network layer design in WSNs must consider the power efficiency, data-centric communication, data aggregation, etc. The transport layer helps to maintain the data flow and may be important if WSNs are planned to be accessed through the Internet or other external networks. Depending on the sensing tasks, different types of application software can be set up and used on the application layer.

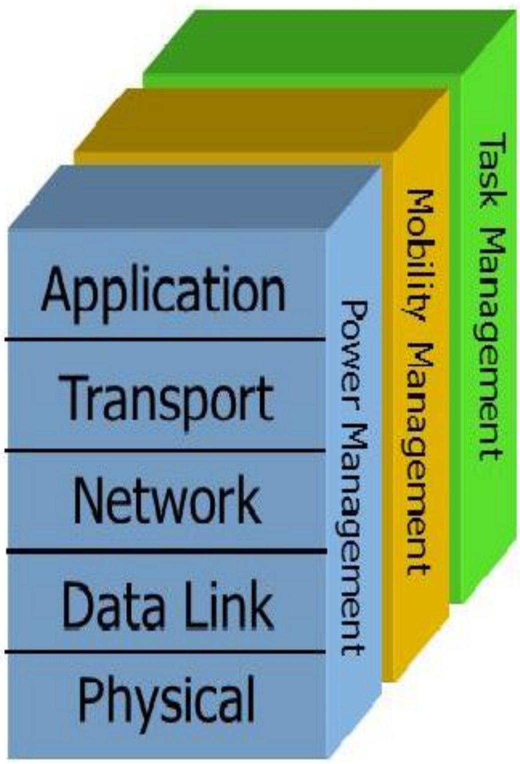

Fig 2.1 Protocol Stack of Wireless Sensor Network

The three main concerns of WSN are:

- **Power Management:** How the sensor nodes use power, e.g. turns off it's circuitry after receiving a message.
- **Mobility Management:** Detects and registers the movement of the sensor nodes.
- **Task Management:** Balances and schedules the sensing tasks given to a specific region.

3 DATA AGGREGATION TECHNIQUES

Data aggregation plays an important role in wireless sensor networks (WSNs) as far as it reduces power consumption and boosts the scalability of the network, especially in topologies that are prone to bottlenecks (e.g. cluster-trees). Some of the previous techniques and briefed below.

3.1 Centralized Approach

This is an address centric approach where each node sends data to a central node via the shortest possible route using a multi-hop wireless protocol. The sensor nodes simply send the data packets to a leader, which is the powerful node. The leader aggregates the data which can be queried.

3.2 In-network Aggregation

In-network aggregation is the Global process of gathering and routing information through a multi-hop network, processing data at intermediate nodes with the objective of reducing resource consumption (in particular energy), and thereby increasing network lifetime. There are two approaches for in-network aggregation: with size reduction and without size reduction. In-network aggregation with size reduction refers to the process of combining and compressing the data packets received by a node from it's neighbors in order to reduce the packet length to be transmitted or forwarded towards sink. Without size reduction is in-network aggregation without size reduction refers to the process merging data packets received from the neighbors in to a single data packet but without processing the value of data.

3.3 Tree based Approach

In the tree based approach, we aggregate by constructing an aggregation tree, which could be a minimum spanning tree, rooted at sink and source nodes are considered as leaves. Each node has a parent node to forward its data. Flow of data starts from leaves nodes up to the sink and therein the aggregation done by parent nodes. Disadvantages of this approach is that in case of the data packet is lost at the level of tree, the data will be lost not for a single level but for the whole related sub tree as well.

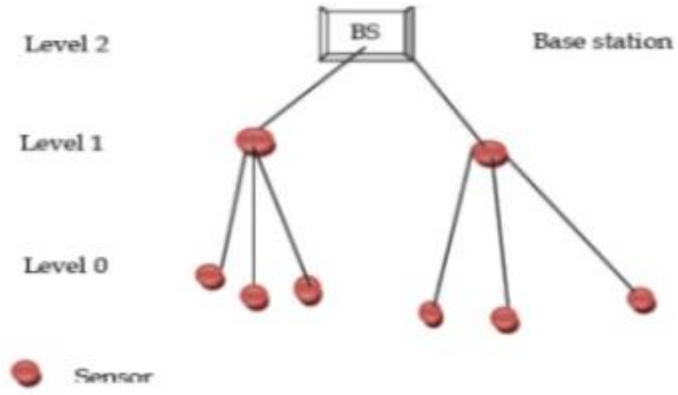

Fig 3.1 Tree based aggregation method for WSN

3.4 Cluster based Approach

In energy-constrained sensor networks of large size, it is inefficient for sensors to transmit the data directly to the sink. In such scenarios, cluster based approach is used. It is a hierarchical approach. In cluster based approach, whole network is divided into several clusters. Each cluster has a cluster head, which is selected among other cluster members. Cluster heads do the role of aggregator which aggregate data received from cluster members locally and transmit to the base station (sink).

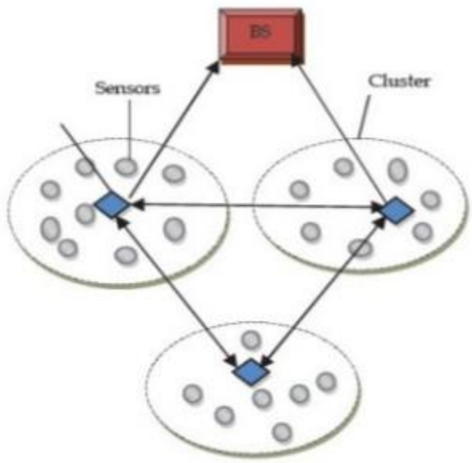

Fig 3.2 Clustered based aggregation method for WSN

4 PROBLEM DEFINITION

Data aggregation protocols aims at eliminating redundant data transmission and thus improve the lifetime of energy constrained WSN. In WSN, data transmission takes place in a multi-hop fashion where each node moves it's data to the neighbouring node closer to the sink. Since closely placed nodes may sense the same data, above approach may not be considered as energy efficient.

An important over the above approach would be clustering where each node sends data to cluster heads and then the cluster head performs aggregation on the received raw data, and then forwards to the sink. Performing aggregation function over cluster head still causes a significant energy wastage.

The data in wireless sensor networks is organized in an efficient manner using data aggregation and data dissemination protocols. Due to the energy constraints in sensor nodes, energy-efficient data aggregation protocols are used to save the node energy and enhance the network life cycle. Deploying additional sensor nodes in the network reduce the resource constraints but increase the rate of data redundancy. This limitation is addressed by the data aggregation protocols in sensor networks. Data aggregation protocols use cluster head node to collect the data, aggregate the data and forward the data to the base station. The primary attributes considered in the design of data aggregation protocols are energy, latency, cluster size and data rate. In this article, we present a novel approach to classify the energy-efficient data aggregation protocols based on structure, search-based and time-based approaches. Analysis for structure-free, structure-based, distance and time-based data aggregation protocols are given in detail. Simulation results indicate that the energy and throughput rate are improved in the cluster-based data aggregation protocols as compared to the structure-free, time-based or search-based data aggregation protocols.

5 PROPOSED APPROACH

The data aggregation function works after the collection of data from the individual sensor nodes. The aggregated data is then sent to the sink and from there the sink sends it to the base station. The flow diagram of the process is given below:

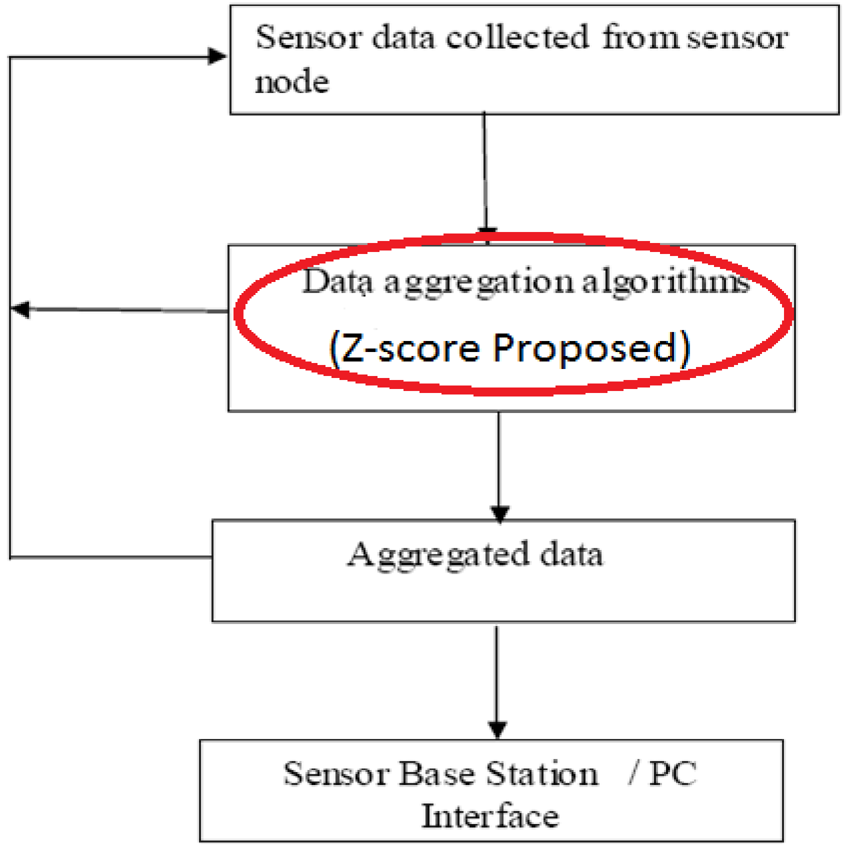

Fig 5.1 Diagrammatic representation of Z score aggregation method for WSN

We have proposed to aggregate the data using the concept of "Z Scores". A Z-score is simply a standardized value which gives us the distance between the raw data (or score) and the mean of the populated data in terms of standard deviation. The Z score for a particular raw score 'x' is defined as:

$$z = \frac{x - \mu}{\sigma}$$

Here:

Z : The Z score obtained for the raw data 'x'
X: The raw data
µ: Mean of the populated data available
σ : The standard deviation of the populated data.

Now, lets look into the procedure followed in each step of our approach:

- Before data aggregation is performed, the sensor network is deployed using the cluster head approach. Each cluster head is assigned such that the nodes which are reporting to that particular cluster head have similar set of data. This can be done by setting the cluster based on the distances between them

- Each cluster head will now collect the data from the individual sensor nodes and store the data in the form of a 2D matrix M[i,j]. Here:

    - i denotes the id of the sensor node from which the cluster head has received the raw data
    - j denotes the number of iterations that have taken place so far.

  So, M[2,3] would have the raw data as : "The data sent by the 2nd node in the 3rd iteration to the cluster head.

- Now, each cluster head will perform the Z Score operation on the matrix elements to perform the necessary data aggregation.

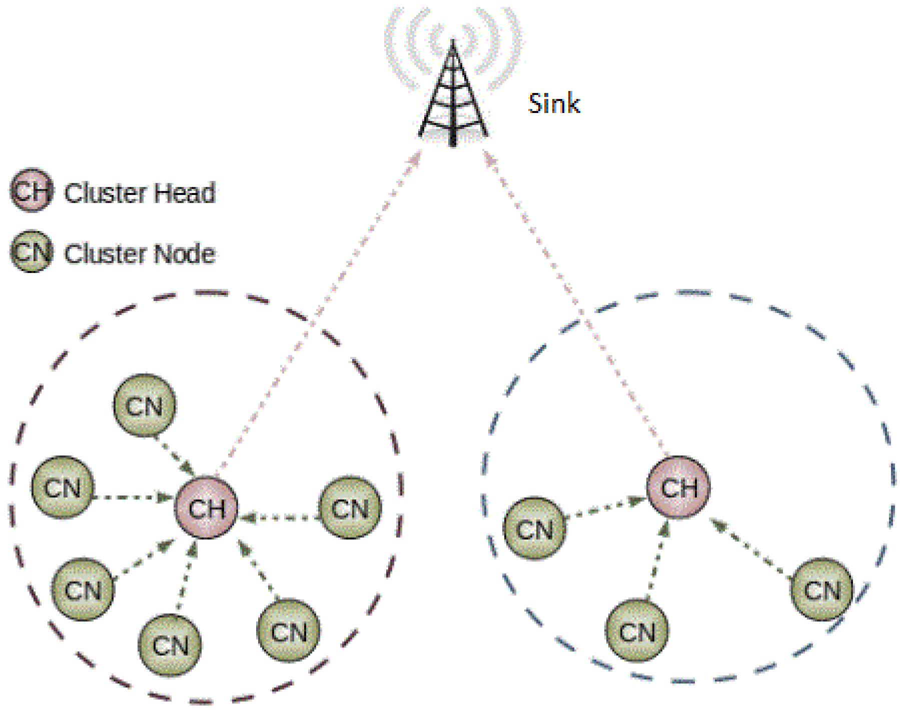

Fig 5.2 The cluster heads sending data to the sink

- Before going into the operations done at the cluster head, lets look at some characteristics of Z scores.

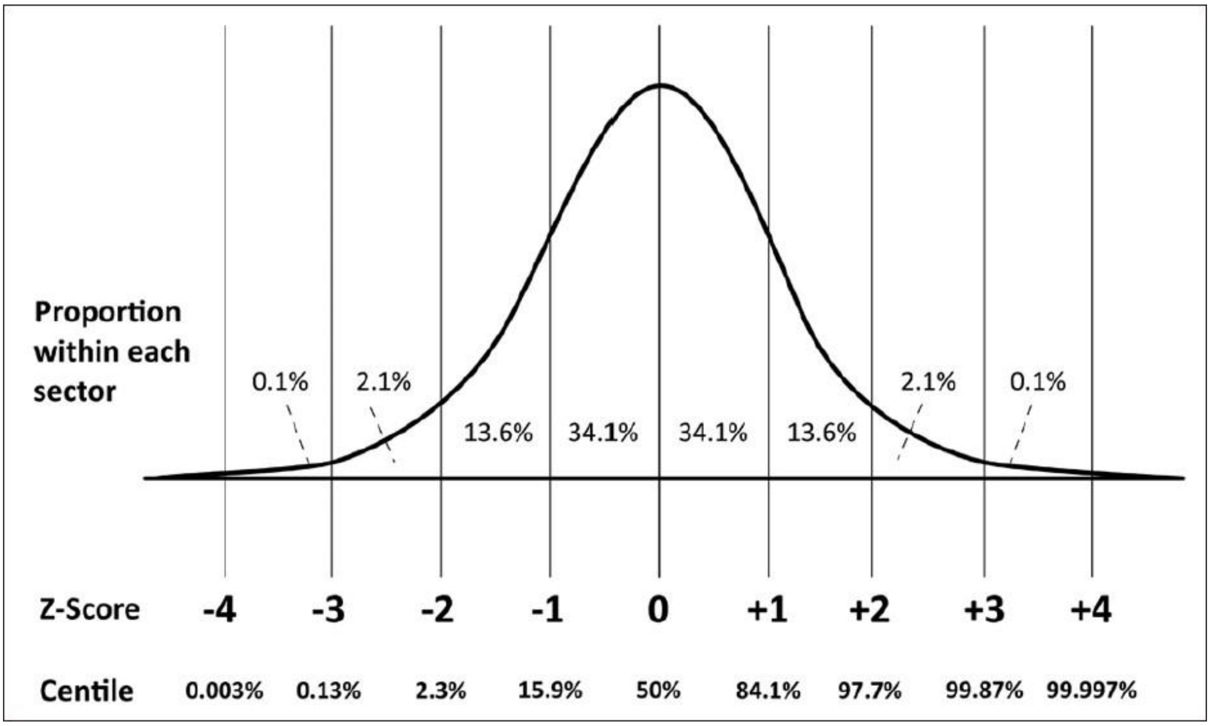

Fig 5.3 Z score distribution curve

As it is clear from the above graph, Z Score follows the normal Bell curve distribution. It can be seen that for a set of populated data, most of the readings will have a Z Score in the range (-2,2). It contains about 95.4% of the data. So, it can be said that any element having a Z Score of above abs(2) is having a value which rather deviates from the main set of values. Using this property of Z Scores, we can define two cases that can occur for an element in the matrix.

<u>Case 1</u>: The Z Score value for a particular element is above abs(2)

This sort of data can be considered as extreme since the probability of encountering such a type of data is about 0.04. So on encountering such an element, the element is directly sent to the sink node. The rest of the elements in the column are considered after this. It is to be noted that the Z Score operation is a column major operation i.e, the elements that are considered are with respect to the values obtained after one iteration.

<u>Case 2:</u> The Z-Score value for a particular element is less than abs(2)

This sort of data elements can be considered as common. So sending such data to the sink would add to redundancy. So on encountering such data, it is left as it is. After the completion of computation of all the elements in one column, the elements which come under this case are considered, their mean value is taken and then just the mean value is sent to the sink.

5.1 Implementation

For implementation of our algorithm, we have used a probabilistic method to determine the cluster heads and used Z-score as a data aggregation technique. Nodes that have been cluster heads cannot become cluster heads again for *P* rounds, where *P* is the desired percentage of cluster heads. Thereafter, each node has a $1/P$ probability of becoming a cluster head again. At the end of each round, each node that is not a cluster head selects the closest cluster head and joins that cluster. The cluster head then creates a schedule for each node in its cluster to transmit its data.

$$T(n) = \frac{P}{1 - P \times (r \bmod P^{-1})} \qquad \forall n \in G$$

$$T(n) = 0 \qquad \forall n \notin G$$

Where n is a random number between 0 and 1
P is the cluster-head probability and
G is the set of nodes that weren't cluster-heads the previous rounds

- If n < T(n), then that node becomes a cluster-head
- The algorithm is designed so that each node becomes a cluster-head at least once

The algorithm is cluster based, and there is random cluster head selection in each round with rotation. The cluster membership is adaptive, and data aggregation takes place at cluster head. There are several desirable properties for protocols on these networks:

- Use 100's - 1000's of nodes
- Maximize system lifetime
- Maximize network coverage
- Use uniform, battery-operated nodes

5.2 Comparison between Z-Score with LEACH and CAG

We have compared our method with a standard TDMA-based MAC protocol, LEACH or Low-Energy Adaptive Clustering Hierarchy, which is a hierarchical protocol in which most nodes transmit to cluster heads, and the cluster heads aggregate and compress the data and forward it to the base station (sink). Each node uses a stochastic algorithm at each round to determine whether it will become a cluster head in this round. LEACH assumes that each node has a radio powerful enough to directly reach the base station or the nearest cluster head, but that using this radio at full power all the time would waste energy.

All nodes that are not cluster heads only communicate with the cluster head in a TDMA fashion, according to the schedule created by the cluster head. They do so using the minimum energy needed to reach the cluster head, and only need to keep their radios on during their time slot.

LEACH also uses CDMA so that each cluster uses a different set of CDMA codes, to minimize interference between clusters.

LEACH includes distributed cluster formation, local processing to reduce global communication, and randomized rotation of the cluster-heads. Together, these features allow LEACH to achieve the desired properties. Initial simulations show that LEACH is an energy-efficient protocol that extends system lifetime.

We have also compared with another popular technique Cluster AGgregation Technique or CAG which is based on data similarity. Clustered AGgregation algorithm forms clusters of nodes sensing similar values within a given threshold (spatial correlation), and these clusters remain unchanged as long as the sensor values stay within a threshold over time (temporal correlation). With CAG, only one sensor reading per cluster is transmitted. Thus, CAG provides energy efficient and approximate aggregation results with small and often negligible and bounded error.

5.3 Results

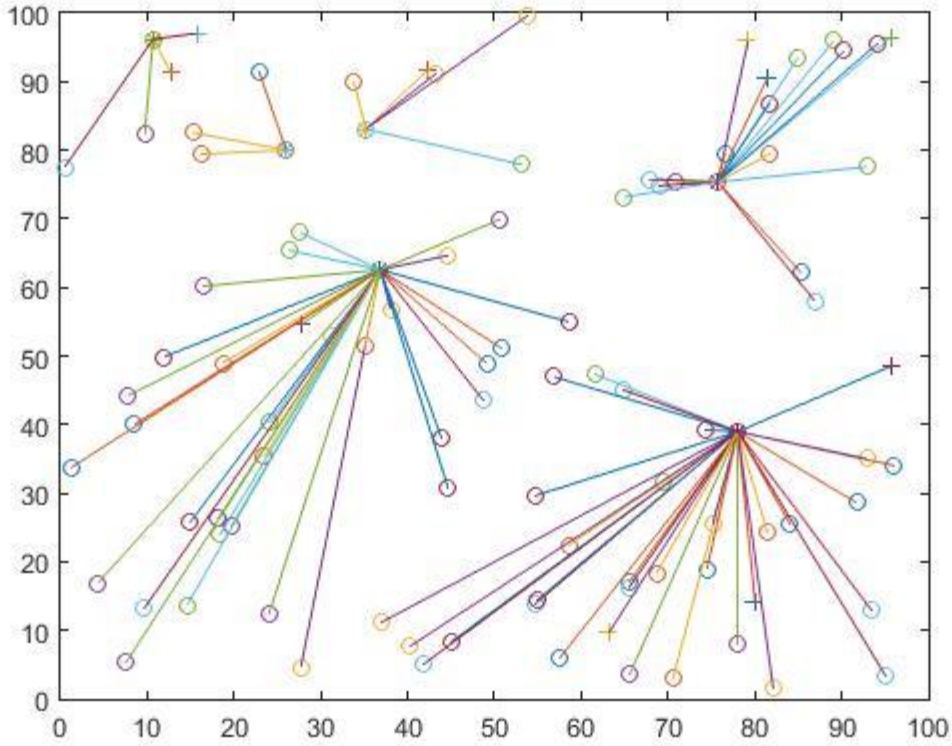

Fig 5.4: Cluster formation using protocol Z score/LEACH – Instance 1

To implement the Z-score technique, we have used a randomised probabilistic method of clustering which is similar to multi-hop LEACH protocol, which chooses the cluster head randomly, and thus the clusters formed are different for every rotation. Moreover, the cluster formation is done with respect to distance after the randomized selection of cluster head.

Fig 5.4 and 5.5 shows how separate clusters are formed with every rotation of Z-score.

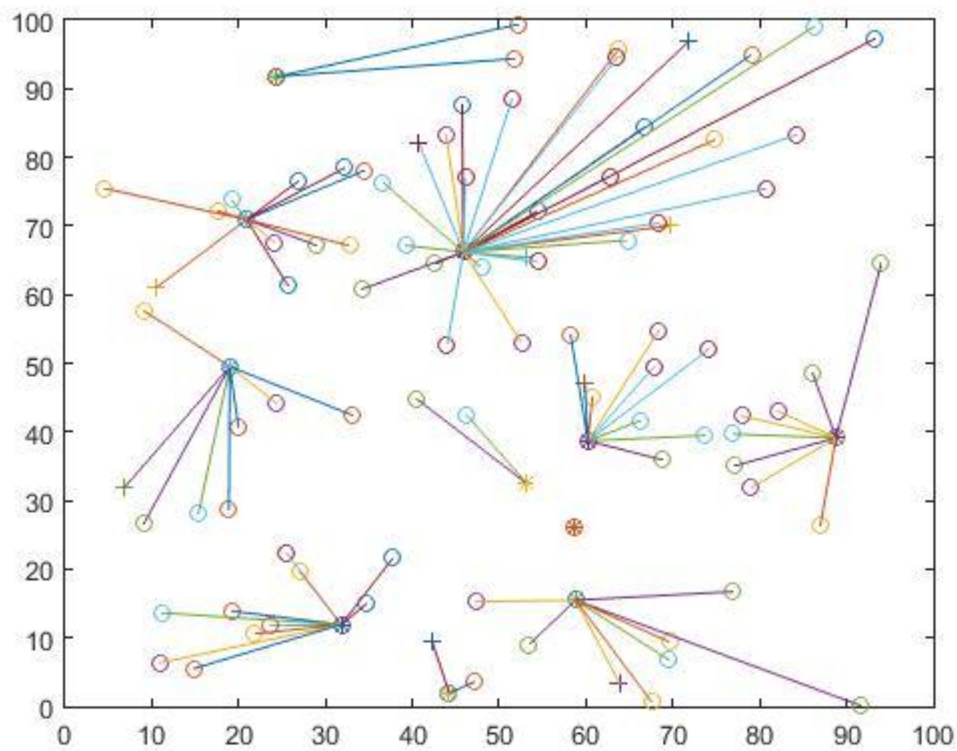

Fig 5.5: Cluster formation using Z score/LEACH – Instance 2

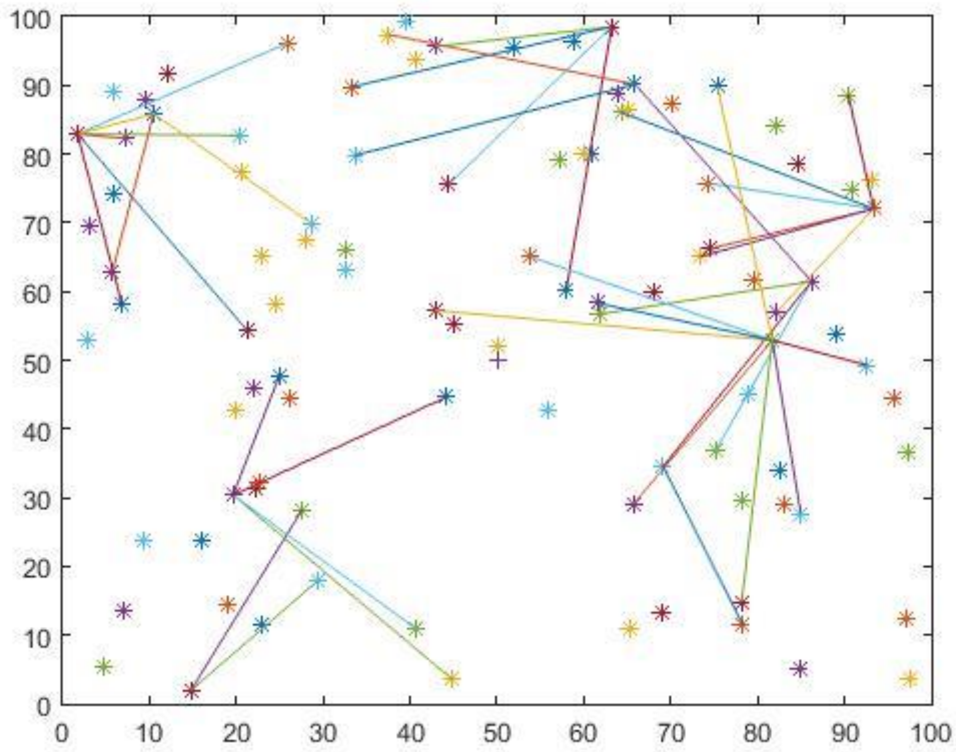

Fig 5.6: Cluster formation by CAG technique

The clusters formed by CAG are based on data similarity approach. We have assumed that the data range is from 0 to 500. We have selected the threshold value to be 20. Hence, Fig 5.6 shows the clustering formed with the given threshold value. The CAG which have been implemented is for sparse networks.

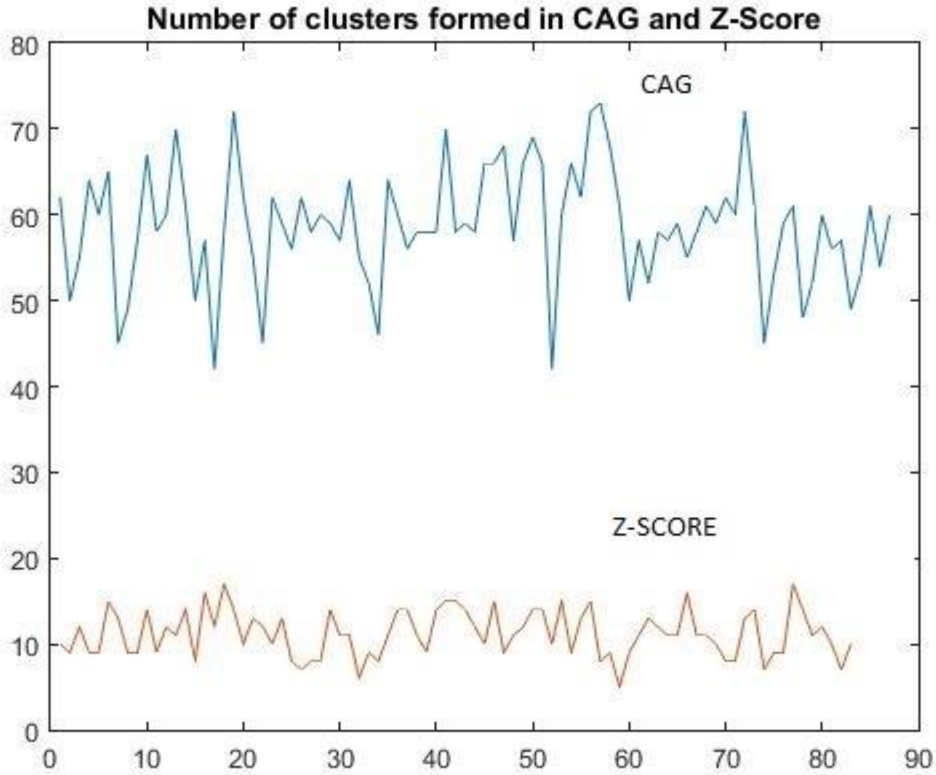

Fig 5.7: Comparison of number of clusters formed by CAG and Z-score.

We also see that the number of cluster formed with Z-score is 3 times less than that of number of clusters formed by CAG. The desired probability for cluster head selection in our approach for LEACH is 0.1. In CAG, as the clusters are formed due to special correlation and there being no upper bound of the clusters to be formed, we find that the number of clusters formed by CAG is much more than when compared with LEACH.

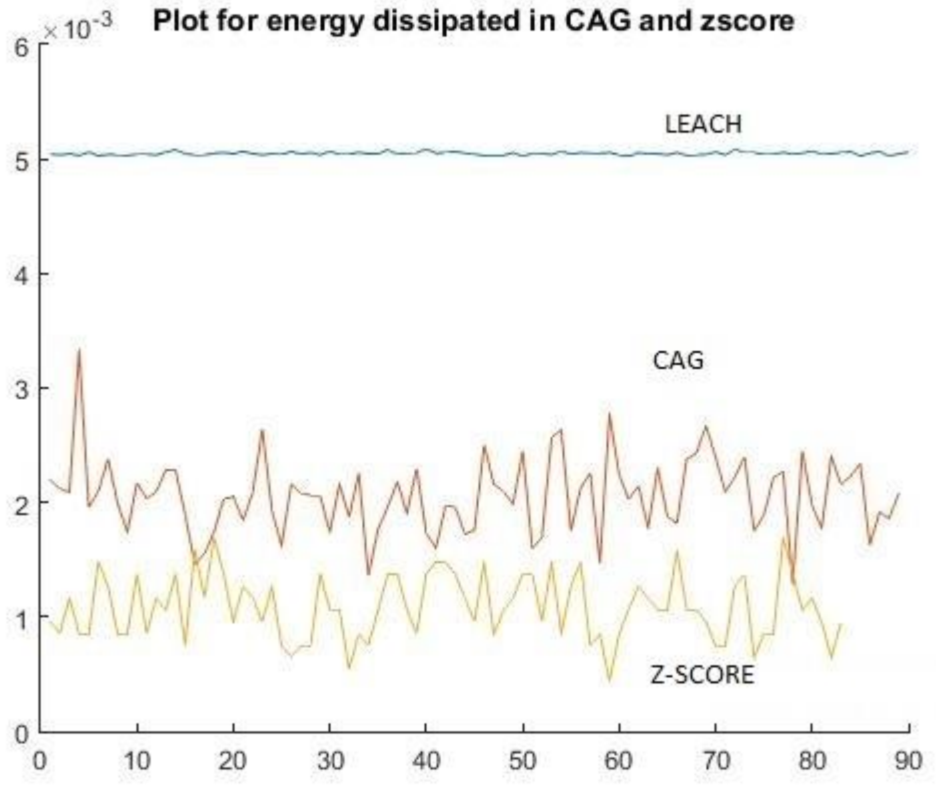

Fig 5.8: Comparison of energy dissipated in transmission by LEACH, CAG Z-score.

We see from the above diagram that energy dissipated by using Z-score is improving LEACH protocol by 4 times. Moreover, when Z-score is applied, it is seen that the energy dissipated is improved by 20-30% when compared with the CAG approach.

5.4 Application:

Sensor network can be deployed for the working in many fields, including in health care for detection of heart attack, for finding environmental variations, for military purposes, etc.

We have tried our approach for environmental temperature variation.

Assumption: The area selected is remote, with temperature variation from 0°C to 50°C. We have assumed the threshold value for data similarity to be 5°C, i.e., if the difference in temperature between two distinct sensor nodes is less than 5°C, then CAG will consider them to belong to the same cluster.

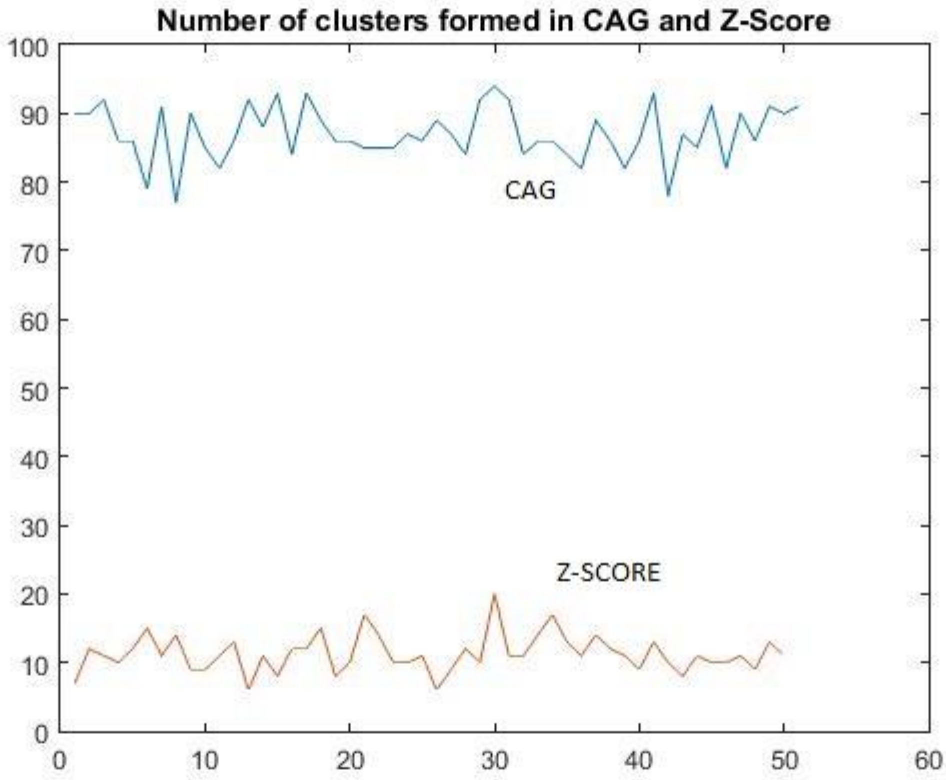

Fig 5.9: Comparison of number of clusters formed by CAG and Z-score.

We see that the number of cluster formed is approximately 9 times more than that of the number of clusters formed by Z-score. As each cluster head will send a separate value, hence the total energy required to send the value from the cluster head to the sink will be more than that of the value for data transmission by Z-score.

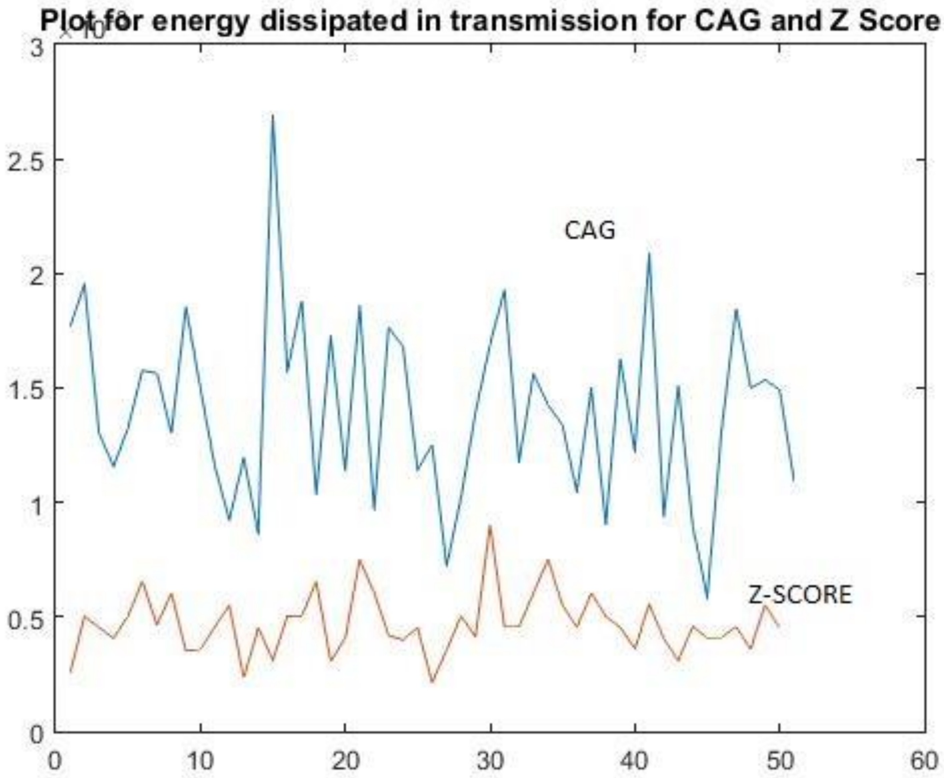

Fig 5.10: Comparison of energy dissipated in transmission by CAG and Z-score.

The energy dissipated by CAG is more than that of Z-Score.

Hence we see our Z-score approach could be used for various different application. Z-score can be used to maximize network coverage, as well as maximize system lifetime.

5.5 Advantages of this approach:

- It will reduce the data traffic considerably. From each cluster in most of the cases, only one packet of data will be forwarded to the sink at each iteration since the probability associated with the encountering of an "extreme" data is just 0.04%

- The redundant data collected at the cluster head will be identified efficiently by the concept of Z Scores and thus will not be sent to the sink. Thus this approach is expected to handle redundancy of data in an effective manner.

- The amount of energy required at the sensor nodes for transmitting data is much more when compared to processing of data. Since this approach tries to minimize the transmission of data as much as possible, therefore it can be considered to be an energy efficient approach.

6 CONCLUSION

Sensor networks have great future prospects and it is important to optimize the amount of energy that is dissipated. Data aggregation is one of those techniques where we can optimize the energy. In this work we have studied the one of the most important parts of data communication in sensor networks- data aggregation, and realized how communication in sensor networks is different from other wireless networks. Wireless sensor networks are energy constrained network. Since most of the energy consumed for transmitting and receiving data, the process of data aggregation becomes an important issue and optimization is needed. Efficient data aggregations not only provide energy conservation but also remove redundancy data and hence provide useful data only.

We see that our Z-score approach can be used as an improvement to LEACH technique, and also its performance is better than other data aggregation technique like CAG. Even the number of clusters formed by Z-score can be defined by probabilistic approach. Hence as the number of cluster formed also influence the energy dissipated, controlling the number of cluster heads keep the energy dissipated in limit. Hence Z-score can be used as a data aggregation technique.

References

[1] Topic title. **http://www.iiests.ac.in**, 2017.

[2] Jennifer Yick, Biswanath Mukherjee, Dipak Ghosal, Wireless Sensor Network Survey, Elsevier, 2008.

[3] BINARYCSE, http://www.binarycse.com/

[4] Mou Wua, Liansheng Tana, Naixue Xiongb, Data prediction, compression and recovery in clustered wireless sensor networks for environment monitoring applications. 2015

[5] R Szewyzck, E Osterweil, J Polastre, M Hamilton, A Mainwaring, D Estrin, Habitat monitoring with sensor networks. 2004