ISSN 2250-3153

Energy Aware Data Aggregation Technique in WSN

Nagaveni.B.Sangolgi*, Syed Khaja Ahmeduddin Zakir**

[*] Department of E&C, Shadan College of Engineering & Technology, Hyderabad
[**] Associate Prof, Department of E&C, Shadan College of Engineering & Technology, Hyderabad

Abstract- Wireless sensor networks consist of sensor nodes with sensing, communicating, computing & storage with battery capacity. Data aggregation is a process or scheme to eliminate redundant transmission & provide fused information to base station which improves energy efficiency & network .lifetime of energy constrained WSN**..**In this paper, we present a survey of data aggregation schemes in Flat & Hierarchical wireless sensor networks & compare them on the basis of metrices such as lifetime, latency and data accuracy. Our main focus on data aggregation in cluster based network using LEACH protocol which delivers 10 times more data than the minimum energy transmission routing which improves system lifetime & reliability of data transmission & energy consumption by a factor of 8 compared to direct transmission. This paper highlights some of the drawbacks and issues in LEACH & the proposed protocol LEACH Acess Point(LEACH-AP).

Index Terms- Clusterhead, Energy Efficiency, LEACH, LEACH-AP

I. Introduction

Wireless sensor networks (WSNs) have been used for numerous applications including military surveillance, facility monitoring and environmental monitoring. Typically WSNs have a large number of sensor nodes with the ability to communicate among themselves and also to an external sink or a base-station [1, 2]. The sensors could be scattered randomly in harsh environments such as a battlefield or deterministically placed at specified locations. The sensors coordinate among themselves to form a communication network such as a single multi-hop network or a hierarchical organization with several clusters and cluster heads. The sensors periodically sense the data, process it and transmit it to the base station. The frequency of data reporting and the number of sensors which report data usually depends on the specific application.A comprehensive survey on wireless sensor networks is presented in [3].

Data gathering is defined as the systematic collection of sensed data from multiple sensors to be eventually transmitted to the base station for processing. Since sensor nodes are energy constrained, it is inefficient for all the sensors to transmit the data directly to the base station. Data generated from neighboring sensors is often redundant and highly correlated. In addition, the amount of data generated in large sensor networks is usually enormous for the base station to process. Hence, we need methods for combining data into high quality information at the sensors or intermediate nodes which can reduce the number of packets transmitted to the base station resulting in conservation of energy and bandwidth. This can be accomplished by data aggregation. *Data aggregation* is defined as the process of aggregating the data from multiple sensors to eliminate redundant transmission and provide fused information to the base station. Data aggregation usually involves the fusion of data from multiple sensors at intermediate nodes and transmission of the aggregated data to the base station (sink). In the rest of the paper, we use the term data aggregation to denote the process of data gathering with aggregation. We also use the term sink to represent the base station

Data aggregation attempts to collect the most critical data from the sensors and make it available to the sink in an energy efficient manner with minimum data latency. Data latency is important in many applications such as environment monitoring where the freshness of data is also an important factor**.** It is critical to develop energy efficient data aggregation algorithms so that network lifetime is enhanced. There are several factors which determine the energy efficiency of a sensor network such as network architecture, the data aggregation mechanism and the underlying routing protocol. In this paper, we describe the influence of these factors on the energy efficiency of the network in the context of data aggregation. We now present a formal definition of energy efficiency.

Energy Efficiency*:* The functionality of the sensor network should be extended as long as possible. In an ideal data aggregation scheme, each sensor should have expended the same amount of energy in each data gathering round. A data aggregation scheme is energy efficient if it maximizes the functionality of the network. If we assume that all sensors are equally important, we should minimize the energy consumption of each sensor. This idea is captured by the network lifetime which quantifies the energy efficiency of the network.

Network lifetime, data accuracy, and latency are some of the important performance measures of data aggregation algorithms. The definitions of these measures are highly dependent on the desired application. We now present a formal definition of these measures.

Network lifetime*:* Network lifetime is defined as the number of data aggregation rounds till □% of sensors die where □ is specified by the system designer. For instance, in applications where the time that all nodes operate together is vital, lifetime is defined as the number of rounds until the first sensor is drained of its energy.

Data accuracy*:* The definition of data accuracy depends on the specific application for which the sensor network is designed.

Latency*:* Latency is defined as the delay involved in data transmission, routing and data aggregation. It can be measured as the time delay between the data packets received at the sink and the data generated at the source nodes.

The rest of the paper is organized as follows. In Section 2, we categorize different data aggregation protocols based on the

ISSN 2250-3153

network architecture involved in data aggregation. Section 3 describes LEACH Protocol In Section 4 Proposed Protocol LEACH-AP Protocol. Section 5 Simulation Results. Section6 describes Conclusion & futurework.

II. DATA AGGREGATION PROTOCOLS BASED ON NETWORK ARCHITECTURE

The architecture of the sensor network plays a vital role in the performance of different data aggregation protocols. In this section, we survey several data aggregation protocols which have specifically been designed for different network architectures.

2.1 Flat networks

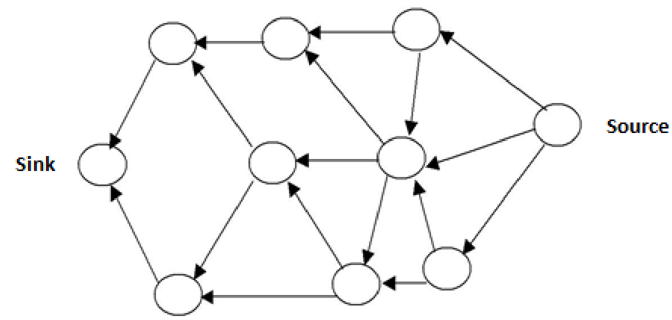
In flat networks, each sensor node plays the same role and is equipped with approximately the same battery power. In such networks, data aggregation is accomplished by data centric routing where the sink usually transmits a query message to the sensors, e.g, via flooding and sensors which have data matching the query send response messages back to the sink. The choice of a particular communication protocol depends on the specific application at hand. In the rest of this subsection, we describe these protocols and highlight their advantages and limitations.

2.1.1 Push diffusion

In the push diffusion scheme, the sources are active participants and initiate the diffusion while the sinks respond to the sources. The sources flood the data when they detect an event while the sinks subscribe to the sources through enforcements. The *sensor protocol for information via negotiation* (SPIN) [4] can be classified as a push based diffusion protocol. The two main features of SPIN are negotiation and resource adaptation. For successful data negotiation, sensor nodes need a descriptor to succinctly describe their observed data. These descriptors are defined in SPIN as *metadata*.

2.1.2 Two phase pull diffusion

Intanagonwiwat et al. [5] have developed an energy efficient data aggregation protocol called directed diffusion. Directed diffusion is a representative approach of two phase pull diffusion. It is a data centric routing scheme which is based on the data acquired at the sensors. The attributes of the data are utilized message in the network. Figure 1 illustrates the interest propagation in directed diffusion. If the attributes of the data generated by the source match the interest, a gradient is set up to identify the data generated by the sensor nodes. The sink initially broadcasts an interest message in the network. The gradient specifies the data rate and the direction in which to send the data. Intermediate nodes are capable of caching and transforming the data. Each node maintains a data cache which keeps track of recently seen data items. After receiving low data rate events, the sink reinforces one particular neighbor in order to attract higher quality data. Thus, directed diffusion is achieved by using data driven local rules.



Impact of source- destination location on directed diffusion

The performance of the data aggregation protocol in directed diffusion is influenced by factors such as the position of source and destination nodes and network topology. Krishnamachari et al. [6] have studied the impact of source-destination placement and communication network density on the energy costs associated with data aggregation. The event radius model (*ER*) and random source (*RS*) model are considered for source placement. In the *ER* model, all sources are assumed to be located within a fixed distance of a randomly chosen " event" location. In the *RS* model, a fixed number of nodes are randomly chosen to be sources.

2.1.3 One phase pull diffusion

Two phase pull diffusion results in large overhead if there are many sources and sinks. Krishnamachari et al. [7] have proposed a one phase pull diffusion scheme which skips the flooding process of directed diffusion. In one phase pull diffusion, sinks send interest messages that propagate through the network establishing gradients. However, the sources do not transmit exploratory data. The sources transmit data only to the lowest latency gradient pertinent to each sink. Hence, the reverse route (from the source to the sink) has the least latency. Removal of exploratory data transmission results in a decrease in control overhead conserving the energy of the sensors.

2.2. Hierarchical networks

A flat network can result in excessive communication and computation burden at the sink node resulting in a faster depletion of its battery power. The death of the sink node breaks down the functionality of the network. Hence, in view of scalability and energy efficiency, several hierarchical data aggregation approaches have been proposed. Hierarchical data aggregation involves data fusion at special nodes, which reduces the number of messages transmitted to the sink. This improves the energy efficiency of the network. In the rest of this subsection, we describe the different hierarchical data aggregation protocols and highlight their main advantages and limitations.

2.2.1 Data aggregation in cluster based networks

In energy constrained sensor networks of large size, it is inefficient for sensors to transmit the data directly to the sink. In such scenarios, sensors can transmit data to a local aggregator or cluster head which aggregates data from all the sensors in its cluster and transmits the concise digest to the sink. This results in significant energy savings for the energy constrained sensors. Figure 2 shows a cluster based sensor network organization. The cluster heads can communicate with the sink directly via long range transmissions or multi hopping through other cluster heads. In this section we discuss three such protocols viz., *Low Energy Adaptive Clustering Hierarchy* (LEACH), *Hybrid Energy Efficient Distributed ClusteringApproach* (HEED) and *clustered diffusion with dynamic data aggregation* (CLUDDA).

Heinzelman [8] et al. were the first to propose an energy conserving cluster formation protocol called LEACH. The LEACH protocol is distributed and sensor nodes organize themselves into clusters for data fusion. A designated node (cluster head) in each cluster transmits the fused data from several sensors in its cluster to the sink. This reduces the amount of information that is transmitted to the sink. The data fusion is performed periodically at the cluster heads. LEACH is suited for applications which involve constant monitoring and periodic data reporting. The two main phases involved in LEACH are: setup phase and steady state phase. The setup phase involves the organization of the network into clusters and the selection of cluster heads.

The steady state phase involves data aggregation at the cluster heads and data transmission to the sink. LEACH employs randomization to rotate cluster heads and achieves a factor of eight improvement compared to the direct approach in terms of energy consumption. LEACH was compared with minimum transmission energy routing (MTE) in which intermediate nodes are chosen such that the sum of squared distances between adjacent nodes of the route is minimized. The simulation results show that LEACH delivers ten times more data than MTE for the same number of node deaths.

Although LEACH improves the system lifetime and data accuracy of the network, the protocol has some limitations. LEACH assumes that all sensors have enough power to reach the sink if needed. In other words, each sensor has the capability to act as a cluster head and perform data fusion. This assumption might not be valid with energy-constrained sensors. LEACH also assumes that nodes have data to send periodically.

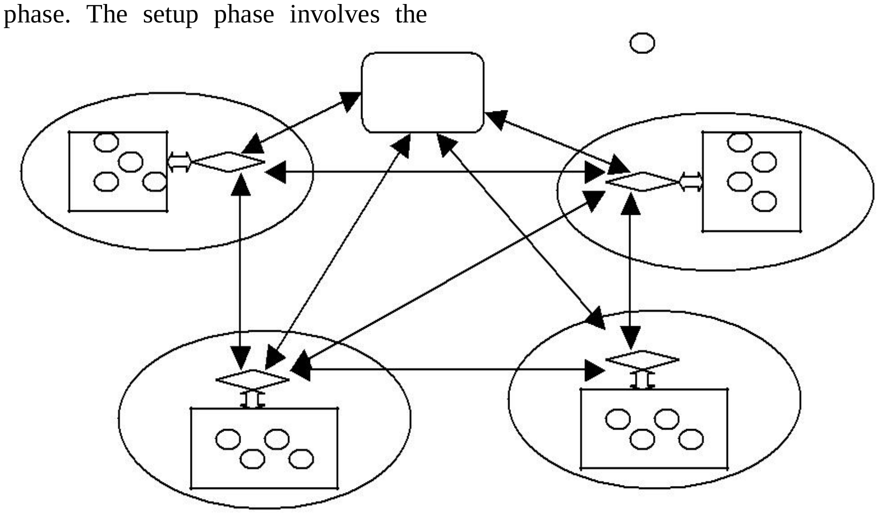


Figure 2: Cluster based sensor network. The arrows indicate wireless communication links.

Younis et al. [10] have proposed HEED whose main goal is to form efficient clusters for maximizing network lifetime. The main assumption in HEED is the availability of multiple power levels at sensor nodes. Cluster head selection is based on a combination of node residual energy of each node and a secondary parameter which depends on the node proximity to its neighbors or node degree. The cost of a cluster head is defined as its average minimum reachability power (AMRP). AMRP is the average of the minimum power levels required by all nodes within the cluster range to reach the cluster head. AMRP provides an estimate of the communication cost.

At every iteration of HEED, each node which has not selected a cluster head, sets its probability P_{CH} of becoming the cluster head as

$$P_{CH} = C \times \frac{E_{residual}}{E_{max}}$$

where *C* denotes the initial percentage of cluster heads (specified by the user) , $E_{residual}$ is the estimated current residual energy of the node and E_{max} is its initial energy corresponding to a fully charged battery. Each node sends a *cluster_head_msg* where the selection status is set to tentative if P_{CH} is less than 1 or final if P_{CH} is 1. A node selects its cluster head as the node with the lowest cost (AMRP) in the set of tentative cluster heads**.** Every node then changes its probability to min(2 $\times$ PCH , 1) in the next iteration. The process repeats until every node is assigned to a cluster head.

HEED improves the network lifetime over gen-LEACH. In gen-LEACH the selection of cluster heads is random which may result in rapid death of certain nodes. However, in HEED the cluster heads are selected such that they are well distributed with minimum communication cost. In addition, the energy dissipated in clustering is less in HEED compared to gen-LEACH. This is due to the fact that gen-LEACH propagates residual energy. To conclude, HEED prolongs network lifetime and achieves a geographically well-distributed set of cluster heads

Recently a hybrid approach [11] has been proposed. The new data aggregation scheme proposed in [11] is called clustered diffusion with dynamic data aggregation (CLUDDA). CLUDDA performs data aggregation in unfamiliar environments by including query definitions within interest messages.

The key idea behind chain based data aggregation is that each sensor transmits only to its closest neighbor. Lindsey et al. [12] presented a chain based data aggregation protocol called power efficient data gathering protocol for sensor information systems (PEGASIS). In PEGASIS, nodes are organized into a linear chain for data aggregation. The nodes can form a chain by employing a greedy algorithm or the sink can determine the chain in a centralized manner. Greedy chain formation assumes that all nodes have global knowledge of the network. The farthest node from the sink initiates chain formation and at each step, the closest neighbor of a node is selected as its successor in the chain.

The PEGASIS protocol has considerable energy savings compared to LEACH.

The main disadvantage of PEGASIS is the necessity of global knowledge of all node positions to pick suitable neighbors and minimize the maximum neighbor distance. In addition, PEGASIS assumes that all sensors are equipped with identical battery power and results in excessive delay for nodes at the end of the chain which are farther away from the leader node.

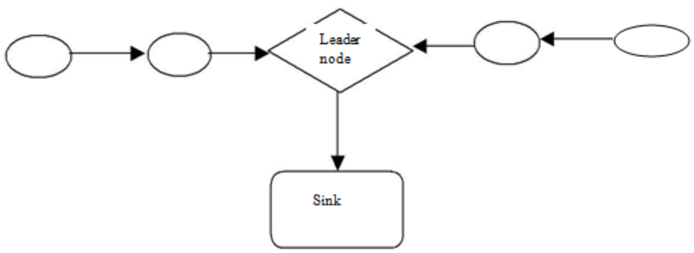


Figure 3: Chain based organization in a 1: Chain based organization in a sensor network.The ovals indicate sensors and the arrows indicate the direction of data transmission.

2.2.3 Tree based data aggregation

In a tree based network, sensor nodes are organized into a tree where data aggregation is performed at intermediate nodes along the tree and a concise representation of the data is transmitted to the root node. Tree based data aggregation is suitable for applications which involve in-network data aggregation. An example application is radiation level monitoring in a nuclear plant where the maximum value provides the most useful information for the safety of the plant. One of the main aspects of tree-based networks is the construction of an energy efficient data aggregation tree.

2.2.4 Grid based data aggregation

Vaidhyanathan et al. [16] have proposed two data aggregation schemes which are based on dividing the region monitored by a sensor network into several grids. They are: grid-based data aggregation and in-network data aggregation. In grid-based data aggregation, a set of sensors is assigned as data aggregators in fixed regions of the sensor network. The sensors in a particular grid transmit the data directly to the data aggregator of that grid. Hence, the sensors within a grid do not communicate with each other.

Table 1: Summary of hierarchical data aggregation protocols

Protocol	**Organization Type**	**Objectives**	**Characteristics**
LEACH	Cluster	Network lifetime: number of nodes that are alive, latency	Randomized cluster head rotation, non-uniform energy drainage across different sensors.
HEED	Cluster	Lifetime: number of rounds until the first node death	Assumption: Multiple power levels in sensors. Cluster heads are well distributed. Achieves better performance than LEACH
PEGASIS	Chain	Lifetime: average energy expended by a node	Global knowledge of the network is required. Considerable energy savings compared to LEACH.
Hierarchical chain based protocols	Chain	Energy□ delay	Binary chain based scheme is eight times better than LEACH and the three level scheme is 5 times better than PEGASIS.
EADAT	Tree	Lifetime: number of	Sink initiated broadcasting

		alive sensors at the end of simulation time	approach. It is not clear how to choose the threshold power (P_{th}) for broadcasting help messages. No comparisons made with other existing aggregation algorithms.
PEDAP-PA	Tree	Lifetime: time until the death of last node	Minimum spanning tree based approach. Achieves two times performance improvement compared to LEACH, PEGASIS.

III. LEACH PROTOCOL

Low Energy Adaptive Clustering Hierarchy (LEACH) is the first hierarchical cluster-based routing protocol for wireless sensor network which partitions the nodes into clusters, in each cluster a dedicated node with extra privileges called Cluster Head (CH) is responsible for creating and manipulating a TDMA (Time division multiple access) schedule and sending aggregated data from nodes to the BS where these data is needed using CDMA (Code division multiple access**)**. Remaining nodes are cluster members. This protocol is divided into rounds; each round consists of two phases:

3.1 Set-up Phase

Each node decides independent of other nodes independent of other nodes if it will become a CH or not. This decision takes into account when the node served as a CH for the last time In the following advertisement phase, the CHs inform their neighborhood with an advertisement packet that they become CHs. Non-CH nodes pick the advertisement packet with the strongest received signal strength.

3.2 Steady-state phase:

Data transmission begins; Nodes send their data during their allocated TDMA slot to the CH. This transmission uses a minimal amount of energy. When all the data has been received, the CH aggregate these data and send it to the BS. LEACH is able to perform local aggregation of data in each cluster to reduce the amount of data that transmitted to the base station.

3.3 LEACH Disadvantages:

Leach is not applicable to networks that are deployed in large region as it uses single hop routing where each node can transmit directly to the cluster head and the sink.

The cluster heads used in the LEACH will consume a large amount of energy if they are located farther away from the sink.
Leach does not guarantee good cluster head distribution and it involves the assumption of uniform energy consumption for the cluster heads.

Leach uses dynamic clustering which results in extra overhead such as the head changes , advertisement that reduces the energy consumption gain.

As the nodes are rotating every time a new cluster head has to be formed ,which consumes energy .

IV. PROPOSED PROTOCOL

In our new version of LEACH protocol, the cluster contains access points which is having very high energy (unlimited) compared to cluster head. So here instead of cluster head ,we are using Access points. The Access points are just like mini base stations. Each and every cluster has an access point. Every time there is no need to form new cluster head as there are no cluster heads. When the nodes rotate then also we have the same access point i.e head . We have implemented this protocol using Ns2 Simulator.

- Network Simulator (Version 2), widely known as NS2, is simply an event driven simulation tool that has proved useful in studying the dynamic nature of communication networks.
- Simulation of wired as well as wireless network functions and protocols (e.g., routing algorithms, TCP, UDP) can be done using NS2.
- In general, NS2 provides users with a way of specifying such network protocols and simulating their corresponding behaviors.
- Due to its flexibility and modular nature, NS2 has gained constant popularity in the networking research community since its birth in 1989.
- NS2 provides users with an executable command ns which takes on input argument, the name of a Tcl simulation scripting file.
- Users are feeding the name of a Tcl simulation script (which sets up a simulation) as an input argument of an NS2 executable command ns.
- In most cases, a simulation trace file is created, and is used to plot graph and/or to create animation.NS2 consists of two key languages: C++ and Object-oriented Tool Command Language (OTcl).
- While the C++ defines the internal mechanism (i.e.,a backend) of the simulation objects, the OTcl sets up simulation by assembling and configuring the objects as well as scheduling discrete events (i.e., a frontend). The C++ and the OTcl are linked together using TclCL.

Red hat Linux operating system

- Company/Developer : Red Hat
- OS family : Unix-like
- Source model : Open source

- Initial release : May 13,1995
- Latest stable release : 9 alias Shirke/March 31,2003
- Kernel type : monolithic kernel type
- License : Various
- Official website : http://www..redhat.com

From the simulation results, we can draw a number of conclusions.

• The delay involved in LEACH-AP is less than the original LEACH.

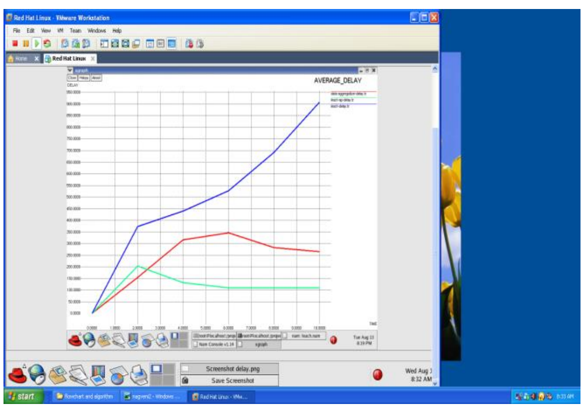
• The energy spent in LEACH-AP is less than the original LEACH.

• The packet delivery ratio in LEACH-AP is more than the original LEACH.

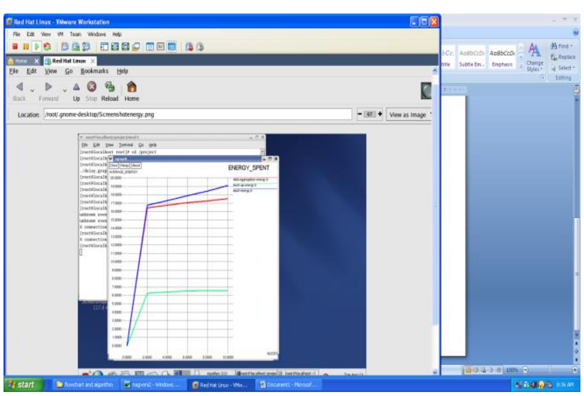
Application of such points is in Military Field or Glacier Monitoring etc where the cluster head has to long last for 6 months or more. So we can go for Access points in such case instead of cluster head which die soon.

V. Simulation Results

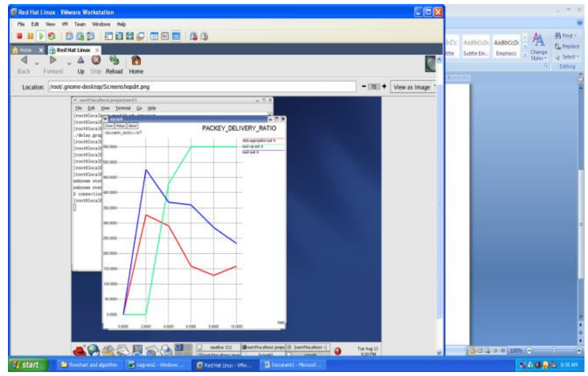
To validate the performance of proposed LEACH-AP protocol, we simulate the protocol and utilize a network with 50 nodes.



Comparison between data aggregation, leach, leach-ap, based on delay



Comparison between data aggregation, leach, leach-ap, based on energy spent



Comparison between data aggregation, leach, leach-ap, based on pdr

VI. Conclusion

In this paper we considered a well known protocol for wireless sensor networks called LEAH protocol which rotate the leader based on available energy, with our new LEACH-AP we have the fixed leader also called access point & conclude that for critical application like military ,glaciour monitoriong our LEACH-AP outperforms in terms of accuracy of data transmission at the cost of abundant of energy source.

References

[1] C. Shen, C. Srisathapornphat, and C. Jaikaeo," Sensor information networking architecture and applications," IEEE Personnel Communications, Aug. 2001, pp.52-59.

[2] S. Tilak, N. Abhu-Gazhaleh, W. R. Heinzelman, " A taxanomy of wireless micro-sensor network models," ACM SIGMOBILE Mobile Comp. Commun. Rev. , vol. 6, no. 2, Apr. 2002, pp. 28-36.

[3] I. Akyldiz, W.Su, Y. Sankarasubramanian and E. Cayirci, " A survey on sensor networks," IEEE Commun. Mag., vol. 40, no. 8, Aug. 2002,102-14.

[4] J. Kulik, W.R. Heinzelman and H. Balakrishnan, " Negotiation-based protocols for disseminating information in wireless sensor networks," Wireless Networks, vol. 8, March 2002, pp. 169-185.

[5] C. Intanagonwiwat, R. Govindan and D. Estrin, " Directed Diffusion: A Scalable and robust communication paradigm for sensor networks," Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking (MobiCOM '00), August 2000.

[6] B. Krishnamachari, D. Estrin, and S. Wicker, " The impact of data aggregation in wireless sensor networks," Proc. 22nd International Conference on Distributed Computing Systems Workshops, July 2002, pp. 575-78.

[7] B.Krishnamachari and J. Heidemann, " Application specific modeling of information routing in wireless sensor networks" , Proc. IEEE international performance, computing and communications conference, vol. 23, pp. 717-722, 2004.

[8] W. R. Heinzelman, " Application-specific protocol architectures for wireless networks" , PhD Thesis, Massachusetts Institute of Technology, June 2000.

[9] W.R. Heinzelman, A.P. Chandrakasan, and H. Balakrishnan, " An application-specific protocol architecture for wireless microsensor networks," IEEE Trans Wireless Communications, October 2002, pp. 660-670.

[10] O. Younis and S. Fahmy, " HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," IEEE Transactions on Mobile Computing, vol .3 , no. 4, Dec 2004, pp. 366-79.

[11] S. Chatterjea and P.Havinga, " A Dynamic data aggregation scheme for wireless sensor networks," Proc. Program for Research on Integrated Systems and Circuits, Veldhoven, The Netherlands, Nov. 2003.

ISSN 2250-3153

[12] S. Lindsey, C. Raghavendra, and K.M. Sivalingam, " Data gathering algorithms in sensor networks using energy metrics," IEEE Trans. Parallel and Distributed Systems, vol. 13, no. 9, September 2002, pp. 924-935.

[13] K. Du, J. Wu and D. Zhou, " Chain-based protocols for data broadcasting and gathering in sensor networks," International Parallel and Distributed Processing Symposium, April 2003.

[14] M. Ding, X. Cheng and G. Xue, " Aggregation tree construction in sensor networks," 2003 IEEE 58th Vehicular Technology Conference, vol.4, no.4, October 2003, pp 2168-2172.

[15] H. O. Tan and I. Korpeoglu, " Power efficient data gathering and aggregation in wireless sensor networks," SIGMOD Record, vol. 32, no. 4, December 2003, pp 66-71.

[16] K. Vaidhyanathan, S. Sur, S. Narravula, P. Sinha, " Data aggregation techniques sensor networks," Technical Report, OSU-CISRC-11/04-TR60, Ohio State University, 2004.

AUTHORS

First Author – Nagaveni.B.Sangolgi, Department of E&C, Shadan College of Engineering & Technology, Hyderabad, nagavenimk289@gmail.com

Second Author – Syed Khaja Ahmeduddin Zakir, Associate Prof., Department of E&C, Shadan College of Engineering & Technology ,Hyderabad
SCETproject12.13@gmail.com