ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 4, Issue 4, pp: (22-27), Month: October - December 2016, Available at: www.researchpublish.com

Efficient Data Collection Using Cluster Technique to Avoid Data Redundancy

M.Senthamilselvi[1], N.Vijayalakshmi[2]

[1]Research Scholar, [2] Associate Professor, [1, 2] Shrimati Indira Gandhi College, Tiruchirappalli, India-620 002

Abstract: **In this paper, a three-layer framework is proposed for mobile data collection in wireless sensor networks, which includes the sensor layer, cluster head layer, and mobile collector (called SenCar) layer. The framework employs distributed load balanced clustering and dual data uploading, which is referred to as LBC-DDU. The objective is to achieve good scalability, long network lifetime and low data collection latency. At the sensor layer, a distributed load balanced clustering (LBC) algorithm is proposed for sensors to self-organize themselves into clusters. In contrast to existing clustering methods, our scheme generates multiple cluster heads in each cluster to balance the work load and facilitate dual data uploading. At the cluster head layer, the inter-cluster transmission range is carefully chosen to guarantee the connectivity among the clusters. Multiple cluster heads within a cluster cooperate with each other to perform energy-saving inter-cluster communications. Through inter-cluster transmissions, cluster head information is forwarded to SenCar for its moving trajectory planning. At the mobile collector layer, SenCar is equipped with two antennas, which enables two cluster heads to simultaneously upload data to SenCar in each time by utilizing multi-user multiple-input and multiple-output (MU-MIMO) technique. The trajectory planning for SenCar is optimized to fully utilize dual data uploading capability by properly selecting polling points in each cluster. By visiting each selected polling point, SenCar can efficiently gather data from cluster heads and transport the data to the static data sink. Extensive simulations are conducted to evaluate the effectiveness of the proposed LBC-DDU scheme. The results show that when each cluster has at most two cluster heads, LBC-DDU achieves over 50 percent energy saving per node and 60 percent energy saving on cluster heads comparing with data collection through multi-hop relay to the static data sink, and 20 percent shorter data collection time compared to traditional mobile data gathering.**

Keywords: **Wireless sensor networks (WSNs), data collection, load balanced clustering, dual data uploading, multi-user multiple-input and multiple-output (MU-MIMO), mobility control, polling point.**

I. INTRODUCTION

The proliferation of the implementation for low-cost, low-power, multifunctional sensors has made wireless sensor networks (WSNs) a prominent data collection paradigm for extracting local measures of interests. In such applications, sensors are generally densely deployed and randomly scattered over a sensing field and left unattended after being deployed, which make it difficult to recharge or replace their batteries. After sensors form into autonomous organizations, those sensors near the data sink typically deplete their batteries much faster than others due to more relaying traffic. When sensors around the data sink deplete their energy, network connectivity and coverage may not be guaranteed. Due to these constraints, it is crucial to design an energy-efficient data collection scheme that consumes energy uniformly across the sensing field to achieve long network lifetime. Furthermore, as sensing data in some applications are time-sensitive, data collection may be required to be performed within a specified time frame. Therefore, an efficient, large-scale data collection scheme should aim at good scalability, long network lifetime and low data latency.

Several approaches have been proposed for efficient data collection in the literature, see, for example, based on the focus of these works, we can roughly divide them into three categories. The first category is the enhanced relay routing in which data are relayed among sensors. Besides relaying, some other factors, such as load balance, schedule pattern and data redundancy, are also considered. The second category organizes sensors into clusters and allows cluster heads to take the responsibility for forwarding data to the data sink. Clustering is particularly useful for applications with scalability requirement and is very effective in local data aggregation since it can reduce the collisions and balance load among sensors. The third category is to make use of mobile collectors to take the burden of data routing from sensors. Although these works provide effective solutions to data collection in WSNs, their inefficiencies have been noticed. Specifically, in relay routing schemes, minimizing energy consumption on the forwarding path does not necessarily prolong network lifetime, since some critical sensors on

the path may run out of energy faster than others. In cluster- based schemes, cluster heads will inevitably consume much more energy than other sensors due to handling intra-cluster aggregation and inter-cluster data forwarding. Though using mobile collectors may alleviate non-uniform energy consumption, it may result in unsatisfactory data collection latency.

In many sensor network applications, there may be multiple base stations to which the sensor nodes report. Each base station selects a group of sensors to construct a "local" data gathering tree. We assume that the base stations have no energy constraint. We thus extend the tree construction problem to construct a data gathering forest for a network with multiple base stations. Each base station should construct a tree which does not intersect with trees constructed by other base stations, and the subset of nodes a base station chooses to construct a tree is not fixed. Hence, it is infeasible to run the tree construction algorithm independently at each base station. This is analogous to network clustering, which cannot be executed independently at each cluster head. Moreover, as will be shown in the paper, running the original tree construction algorithm iteratively could result in poor overall performance. Thus, we need to intelligently extend our framework to construct a maximum lifetime data gathering forest.

In the forest case, the implementation is decentralized among trees, but each base station still makes centralized decisions in its local tree. For applications where a powerful base station is unavailable, distributed implementations of these algorithms are needed. We plan to investigate a fully distributed implementation of the algorithms. Third, we will quantitatively study the impact of wireless transmission errors. Fourth, our work assumes that all nodes transmit at the same power and do not dynamically adjust the transmission power level. Data gathering in the case when nodes dynamically adjust their transmission power levels is an open issue that we plan to investigate.

Nodes in a wireless ad hoc network are usually powered by a limited capacity battery. As nodes' batteries are drained and they stop functioning, the wireless ad hoc network will eventually cease to be usable. Informally speaking, we refer to the length of the time that the network operates prior to becoming unusable as the *network lifetime*. A formal definition of network lifetime is not straightforward and may depend on the application scenario in which the network is used. In the literature, network lifetime has often been defined as the time for the first node to die, or as the time for a certain percentage of network nodes to die. Alternatively, network lifetime has been defined in terms of the packet delivery rate or in terms of the number of a live flows, thus accounting for the "quality of communication" the a live nodes achieve.

II. PROBLEM DEFINITION

Relay routing is a simple and effective approach to routing messages to the data sink in a multi-hop fashion. Devised a coordinated transfer schedule by choosing alternate routes to avoid congestions. Studied the construction of a maximum-lifetime data gathering tree by designing an algorithm that starts from an arbitrary tree and iteratively reduces the load on bottleneck nodes. studied deployments of relay nodes to elongate network lifetime. Gnewali et al. evaluated collection tree protocol (CTP) via testbeds . CTP computes wireless routes adaptive to wireless link status and satisfies reliability, robustness, efficiency and hardware independence requirements. However, when some nodes on the critical paths are subject to energy depletion, data collection performance will be deteriorated. Another approach is to allow nodes to form into clusters to reduce the number of relays.Proposed a cluster formation scheme, named LEACH, which results in the smallest expected number of clusters. However, it does not guarantee good

cluster head distribution and assumes uniform energy consumption for cluster heads.Further proposed "HEED," in which a combination of residual energy and cost is considered as the metric in cluster headselection.

Finding the longest common subsequence (LCS) of multiple sequences (the multiple longest common subsequence problem (MLCS)), which corresponds to the measurement of sequence similarity, is a fundamental problem in many fields, including computer science and molecular biology. It is an NP-hard problem and has been studied extensively.

Although many exact algorithms have been proposed, they are not very useful in practice for large-size problems with many sequences due to high time and space complexity. When dealing with large-size data sets in real applications, for those problems proved to be NP-hard it is more important to find a good approximate solution within a reasonable time rather than waiting long for an exact solution.

A. Problem Analysis:

Sensors are generally densely deployed and randomly scattered over a sensing field and left unattended after being deployed, which makes it difficult to recharge or replace their batteries. After sensors form into autonomous organizations, those sensors near the data sink typically deplete their batteries much faster than others due to more relaying traffic. When sensors around the data sink deplete their energy, network connectivity and coverage may not be guaranteed. Due to these constraints, it is crucial to design an energy-efficient data collection scheme that consumes energy uniformly across the sensing field to achieve long network lifetime. Furthermore, as sensing data in some applications are time-sensitive, data collection may be required to be performed within a specified time frame. Therefore, an efficient, large-scale data collection scheme should aim at good scalability, long network lifetime and low data latency.

B. Problem Solution:

In this paper, we propose first, we propose a distributed algorithm to organize sensors into clusters, where each cluster has multiple cluster heads. In contrast to clustering techniques proposed in previous works, our algorithm balances the load of intra-cluster aggregation and enables dual data uploading between multiple cluster heads and the mobile collector. Second, multiple cluster heads within a cluster can collaborate with each other to perform energy-efficient inter cluster transmissions. Different from other hierarchical schemes, in our algorithm, cluster heads do not relay data packets from other clusters, which effectively alleviate the burden of each cluster head. Instead, forwarding paths among clusters are only used to route small-sized identification (ID) information of cluster heads to the mobile collector for optimizing the data collection tour. Third, we deploy a mobile collector with two antennas (called SenCar in this paper) to allow concurrent uploading from two cluster heads by using MU-MIMO communication.

III. PROCESS FLOW

A. Login and File Upload:

Authorized end-user can get login in the login form and then they can access the split and send form in that form they can upload their desired file as a source file of the source processor that file will be split and send to multiple destinations at a time. After complete the file upload process we will calculate the uploaded file size and show the file size in a label and filename in separate label

B. Split and Send:

Source File will be split into number parts based on number of Processor we have as destination, because of to reduce the total communication round trip between the source and multiple destinations Processors. Split File will be Sending to the Corresponding Destination Processors and also getting the acknowledgment from that destinations Using Approximation Algorithm and TCP server socket Programming.

C. Inter Process of Destination Systems:

Each and every destinations having the parts of the original file that is split files, so every processor will send their holding file to rest of the destination processors, this process will take up to all destinations must receive all split files, this process will reduce the total communication round trip between the source processor and our multiple destination processors.

D. File Merge and Save:

While all destination processors received all part of the split files, Then we will merge hat received files into single file that is original File at all destination processors. That original file will be saved into the user desired location of the local disk.

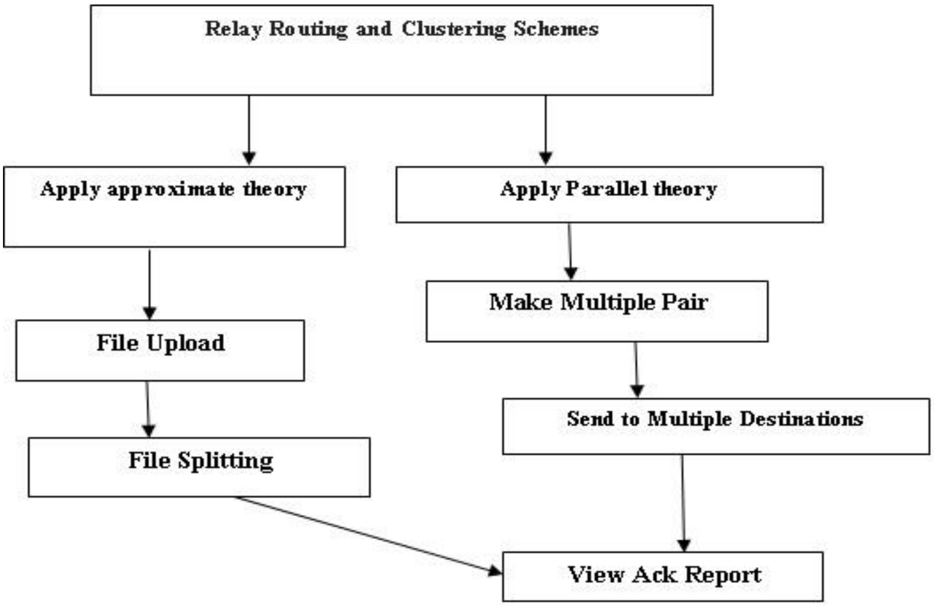

Figure 1 Process Flow

IV. EXPERIMENTAL RESULTS

A. Implementation:

A simple incentive mechanism for P2P systems is the "tit-for-tat" strategy, where peers receive only as much as they contribute. A free rider that does not upload data chunks to other peers cannot get data chunks from them and suffers from poor streaming quality. Due to its simplicity and fairness, this scheme has been adopted by BitTorrent . Though this strategy can increase the cooperation between peers to a certain level, it is shown in literature that it may perform poorly in today's internet environment due to the asymmetry of the upload and download bandwidths.Unlike the "tit-for-tat" strategy, which enforces compulsory contribution from peers, another category of incentive mechanisms stimulate peers to contribute to the system by indirect reciprocity. In these incentive mechanisms, the contribution of each peer is converted to a score which is then used to determine the reputation or rank of the peer among all the peers in the network. Peers with a high reputation are given a certain priority in utilizing the network resources, such as selecting peers or desirable media data chunks. Therefore, peers with a high reputation have more flexibility in choosing desired data suppliers and thus are more likely to receive high-quality streaming.

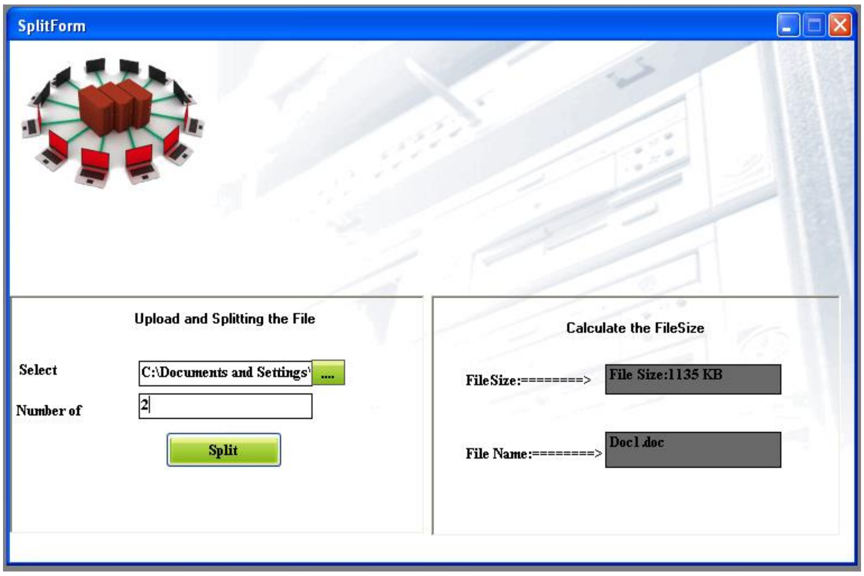

Figure 2 Packet Upload

Figure. 2 show the network setup in this experiment. The authors discussed how to apply game theory to the design of incentive mechanisms for P2P networks at a high level. It is pointed out that straightforward use of results from traditional game theory do not fit well with the requirements of P2P networks. The utility functions must be customized for P2P networks. In a simple, selfish, link-based incentive mechanism for unstructured P2P file sharing systems was proposed.

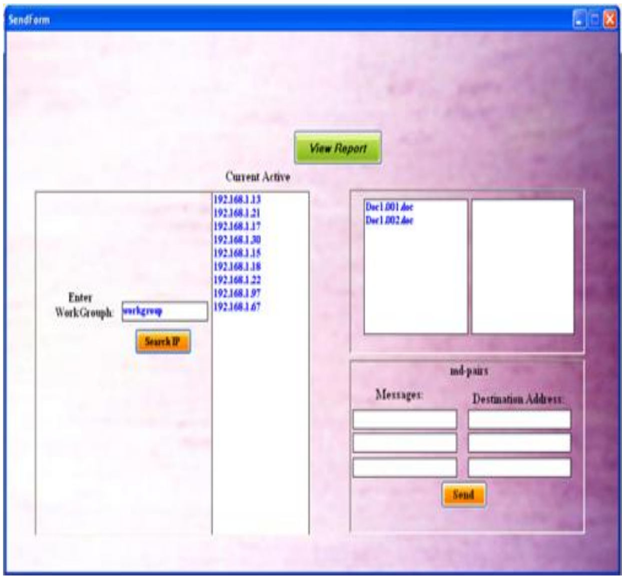

Figure 3 Packet Split

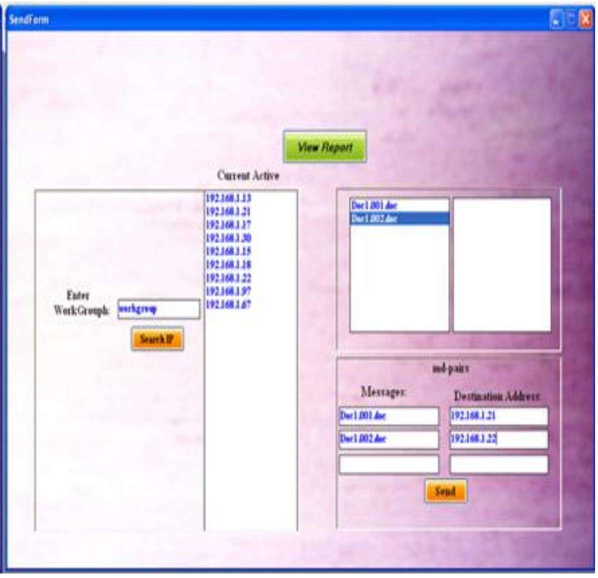

Figure 4 Packet Split

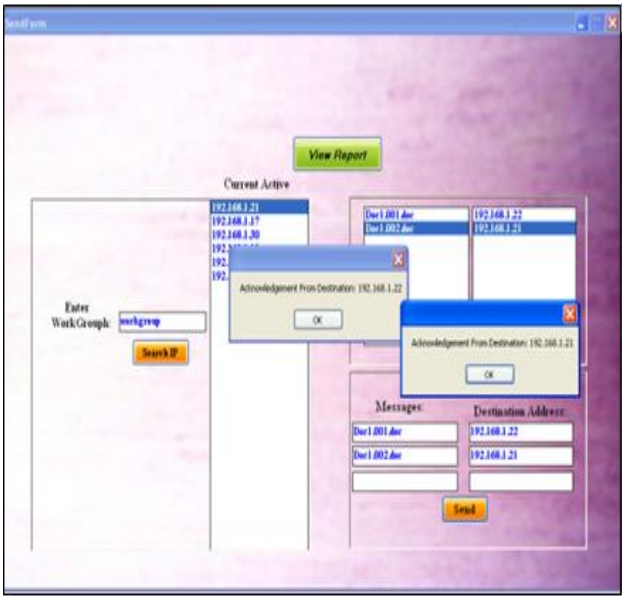

Figure 5 Receive Acknowledgement

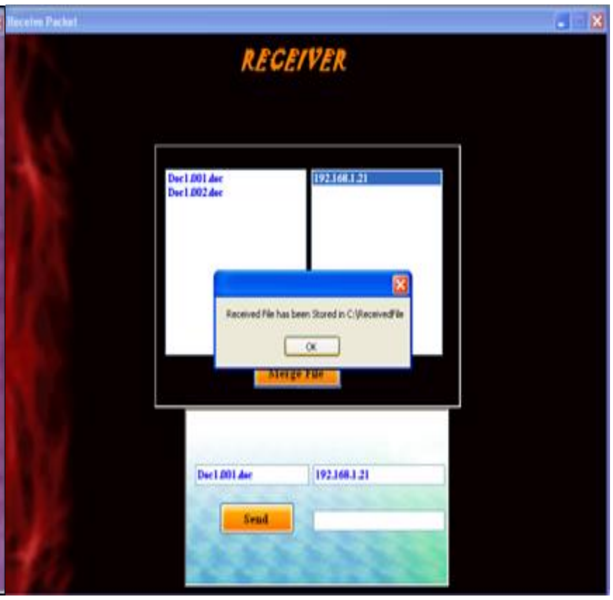

Figure 6 Received File

V. RESULT AND DISCUSSION

Deployments of sensor networks are taking place. Using a controlled mobile element is a promising approach to collect data from these sensor nodes. We showed that as the network scales, using a single mobile element may not be sufficient, and would require multiple of them. The sensor nodes and (or) the mobile elements may not be uniformly placed in practice, necessitating the use of load balancing, so that each mobile element as far as possible, serves the same number of sensor nodes. We gave a load balancing algorithm, and described the mechanism these multiple mobile elements can be used. Finally we presented simulation results justifying our approach. The work presented here can be extended in many directions. For load balancing, we can remove the assumption that each sensor node can talk to at least one mobile element.First, we compare the average energy consumption for each sensor and the maximum energy consumption in the network. The results here are only the average of the connected networks in the experiments.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 4, Issue 4, pp: (22-27), Month: October - December 2016, Available at: www.researchpublish.com

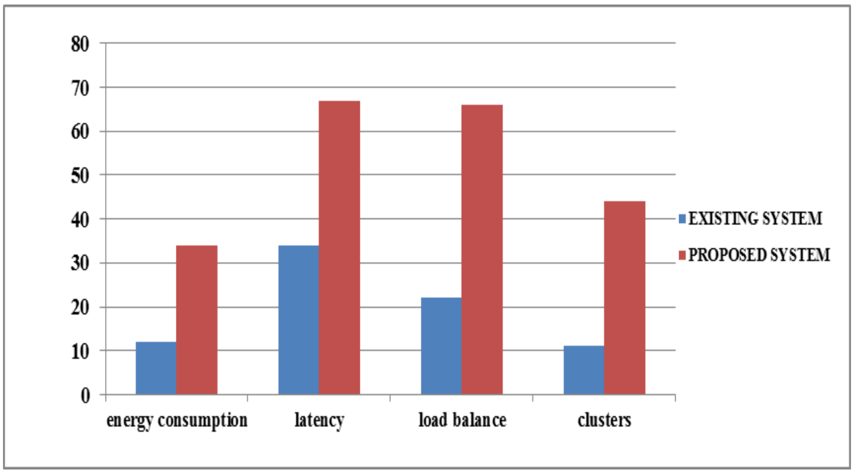

Figure 7 Experimental Result

However, the mobile schemes can work well not only in connected networks but also in disconnected networks, since the mobile collector acts as virtual links to connect the separated subnet works. This is because that though mobility incurs extra moving time, single-hop transmissions for both data aggregation and uploading save time in routing significantly, whereas multichip traffic relay to the static sink may not scale well when the number of nodes increases. Finally, the advantage of mobile MIMO comparing to mobile SISO is observed by saving 20 percent time in total. This is due to the fact of simultaneous data uploading to SenCar at the polling points in the mobile MIMO approach.

VI. CONCLUSION

In this paper, we have proposed the LBC-DDU framework for mobile data collection in a WSN. It consists of sensor layer, cluster head layer and SenCar layer. It employs distributed load balanced clustering for sensor self-organization, adopts collaborative inter-cluster communication for energy-efficient transmissions among CHGs, uses dual data uploading for fast data collection, and optimizes SenCar's mobility to fully enjoy the benefits of MU-MIMO. Our performance study demonstrates the effectiveness of the proposed framework. The results show that LBC-DDU can greatly reduce energy consumptions by alleviating routing burdens on nodes and balancing workload among cluster heads, which achieves 20 percent less data collection time compared to SISO mobile data gathering and over 60 percent energy saving on cluster heads. We have also justified the energy overhead and explored the results with different numbers of cluster heads in the framework.

REFERENCES

[1] Multiple Controlled Mobile Elements (Data Mules) for Data Collection in Sensor Networks,David Jea, Arun Somasundara, and Mani Srivastava Networked and Embedded Systems Laboratory, Department of Electrical Engineering, UCLA

[2] Energy-efficiency of MIMO and Cooperative MIMO Techniques in Sensor Networks , Shuguang Cui, *Student Member, IEEE,* Andrea J. Goldsmith, *Senior Member, IEEE* ,and Ahmad Bahai, *Member, IEEE*

[3] Max-Min D-Cluster Formation in Wireless AdHoc Networks Alan D. Amis Ravi Prakash Thai H.P. Vuong Dung T. Huynh, Department of Computer Science, University of Texas at Dallas, Richardson, Texas 75083-0688 ,E-mail: aamis@telogy.com, ravip@utdallas.edu, tvuong@nortelnetworks.com, ,huynh@utdallas.edu

[4] Space–Time Block Codes from Orthogonal Designs, Vahid Tarokh, *Member, IEEE*, Hamid Jafarkhani, and A. R. Calderbank, *Fellow, IEEE*

[5] Constructing Maximum-Lifetime Data Gathering Forests in Sensor Networks , Yan Wu, Zhoujia Mao, Sonia Fahmy, Ness B. Shroff,E-mail: yanwu@microsoft.com, maoz@ece.osu.edu, fahmy@cs.purdue.edu, shroff@ece.osu.edu

[6] Investigating Upper Bounds on Network Lifetime Extension for Cell-Based Energy Conservation Techniques in Stationary Ad Hoc Networks, Douglas M. Blough, School of Electrical and Computer Engineering,Georgia Institute of Technology ,Atlanta GA 30332-0250 – USA ,doug.blough@ece.gatech.edu.

[7] ZMAC: a Hybrid MAC for Wireless Sensor Networks ,Injong Rhee, Ajit Warrier, Mahesh Aia and Jeongki Min , Dept. of Computer Science, North Carolina State University , Raleigh, NC 27695.