Data Aggregation Protocols in Wireless Sensor Networks

Gunti Spandan[1], Archana Patel [2], C R Manjunath[3] , Nagaraj GS[4]

[1,2] MTech 4th sem, Dept of CSE, SET, Jain University,

[3] Asst prof, Dept of CSE, Jain University,

ABSTRACT

Past few years have witnessed increased interest in the potential use of wireless sensor networks (WSNs) in a wide range of applications and it has become a hot research area. There is need of for fast data access in WSN, as resource-constrained is a critical issue in Wireless Sensor Networks (WSNs) applications. Routing protocols are in charge of discovering and maintaining the routes in the network. Routing protocols with low energy consumption plays a very important role in prolonging the lifetime of sensor network. Owing to a variety of advantages, clustering is becoming an active branch of routing technology in WSNs. Tree-based and Cluster-based routing protocols have proven to be effective in network topology management, energy minimization, data aggregation and so on. Data 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 this paper, various data centric aggregation techniques is discussed like TAG, EADAT, AGIT, SRTSD and PEDAP protocols under tree based and LEACH, PEGASIS, TEEN, APTEEN and HEED under cluster based approach for WSN. Furthermore this paper gives an overview tree-cluster based routing protocols.

KEYWORDS: Wireless sensor network, Tree-based, Cluster-based, routing protocol, Data aggregation, life time, energy minimization;

I. INTRODUCTION

Wireless sensor networks (WSN) are is special kind ad-hoc network, having abilities of sensing, processing and wireless connectivity. Wireless Sensor Network (WSN) contains hundreds or thousands of sensor nodes have the ability to communicate among each other, have limited energy source, energy constrained and bandwidth. 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 to collect the data to sink node in the WSN. Because of limited computing resources of the sensors presents a major challenges for routing protocols and algorithms. Considerable techniques are required to make them energy efficient that would increase the life-time of a WSN [1][2][3][6].Since sensor nodes are energy constrained, it is inefficient for all the sensors to transmit the data directly to the base station. Data Aggregation is the global process of gathering and routing information through a multi hop network with the objective of reducing resource consumption (in particular energy) and prolong the network lifetime in WSNs [4].

II. DATA AGGREGATION

The main purpose of the data aggregation is to reduce the power consumption by minimizing the number of data transmissions. *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. All the aggregation nodes collect data from their children nodes and calculate the aggregation value. Then only the aggregated values are forwarded towards the data sink. The aggregate value may be average, maximum (minimum), summation, etc. which is calculated according to the application requirements. .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. Data aggregation usually involves the fusion of data from multiple sensors at intermediate nodes and transmission of the aggregated data to the base station.

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. The efficiency of data aggregation algorithms depends on the correlation among the data generated by different information sources (sensor units). A correlation can be either *spatial* or *temporal*. *Aggregation gain* is defined as the measure of reduction in the communication traffic due to the aggregation. The most important ingredient for aggregation is a well designed routing protocol, classified as *Classic (address centric) routing protocols* typically forward data along the shortest path to the destination and *Data centric routing protocols* forward data based on the packet content and choose the next hop in order to promote in-network aggregation and minimize energy expenditure [4][5][6].

2.1.Classification of Data Aggregation Mechanisms [7]According the WSN, Data aggregation mechanisms can be classified as structure-free, structure-based and hybrid structure. When sensor nodes are randomly deployed in the environment, by nature, they require a structure-free mechanism. When sensor nodes are deployed at a large scale, it becomes difficult in terms of data aggregation and management the WSNs. whereas the structure-based data aggregations are defined with a set of algorithms, which divides the network into groups and/or levels. This group manages separately their data aggregation and reduced view of the entire network. However the structure-based mechanisms require an extra charge to organize the network and to maintain organization during the network lifetime. Hybrid structure combines characteristics of both structure-free and structure-based is depending upon application.

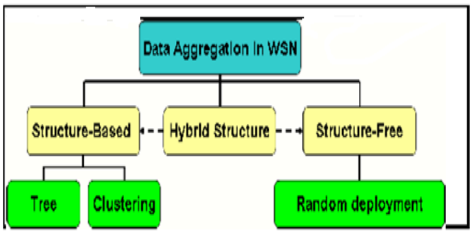

Fig-1 Classification of data aggregation

2.2 Structure-Free Data Aggregation

A WSN is as a multi-hop Ad-Hoc network, where no infrastructure is available to connect the network nodes. The typical communication architecture for a structure-free data aggregation is the basic client/server architecture. To achieve the scalability, it is important to consider several key points such as the high amount of detected data and the communication sessions required to send them to the server. As the sensor nodes are energy constrained, it is inefficient for all the sensors to transmit the data directly to the base station. Because the communication is very expensive in terms of energy compared to the local processing.

2.3 Structure-Based Data Aggregation:

The structures based data aggregation mechanism can be classified as either tree-based or cluster-based.

a) Tree-Based: In a tree-based structure, the sensor nodes are organized into a tree like structure, data aggregation is performed at intermediate nodes along the tree.In the tree-based approach, aggregation is achieved 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 there in the aggregation done by parent nodes shown in figure-2.

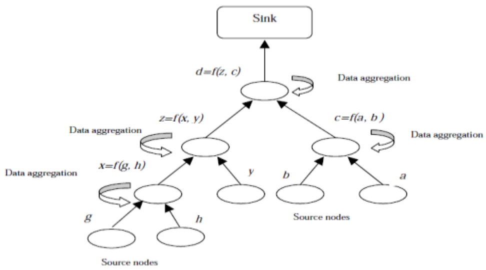

Fig-2

b) Cluster-Based: In a cluster-based structure, the network is partitioned into subgroups. Each subgroup is called cluster. In each cluster, there are several sensor nodes in which a node is assigned as a "cluster head" (CH) as in figure-3. The CHs are designated to send their cluster nodes data to the sink. A CH is usually the master and the sensor nodes are slaves, this master/slave mechanism allows tight traffic control because no node is allowed to transmit outside the cluster, and no communication is allowed between slaves except through the master.

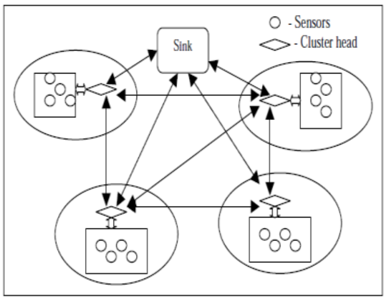

Fig-3

III. DATA AGGREGATION PROTOCOLS

Data aggregation protocols are designed based on network architectures. It is divided into flat, hierarchal and location based protocols. The main objective of data aggregation is to increase the network lifetime by reducing the resource consumption of sensor nodes (such as battery energy and bandwidth). Here, we are mainly focusing on tree-based, cluster-based class of routing protocols [4]. While increasing network lifetime, data aggregation protocols may degrade important quality of service metrics in wireless sensor networks, such as data accuracy, latency, fault-tolerance, and security. Therefore, the design of an efficient data aggregation protocol is an inherently challenging task. Hence, the architecture of the sensor network plays a vital role in the performance of different data aggregation protocols. There are several protocols that allow routing and aggregation of data packets simultaneously [9].

3.1Routing Challenges and Design Issues:

Following are the routing challenges and design issues that affect routing process in WSNs [3, 8, 10]:

- **Node deployment:** Sensor node deployment usually varies with application requirements and affects the performance of routing protocol. The commonly used random or deterministic deployments, some applications may support redeployment of nodes, for instance, to eliminate sensing holes in the network. Inter-sensor communication is normally within short transmission ranges due to energy and bandwidth limitations. Hence, it is most likely that a route will consist of multiple wireless hops.

- **Energy consumption without losing accuracy:** Sensor nodes can use maximum of their limited supply of energy to perform computations and for transmitting the information in a wireless environment. Energy conserving forms of communication and computation are essential, as the Sensor node lifetime is strongly depending on the battery lifetime. In a multihop WSN, each node plays a dual role as data sender as well as data router. The malfunctioning of some sensor nodes due to power failure can cause topological changes and might require rerouting of packets and reorganization of the network.

- **Node/Link Heterogeneity:** Normally all sensor nodes were assumed to be homogeneous, i.e., having equal capacity in terms of computation, communication, and power. However, sensor node can have different role or capability, depending on the application. The existence of this type of sensors (heterogeneous set) raises many technical issues related to data routing. In hierarchical approaches, the protocols designate a cluster head node different from the normal sensors. These cluster heads can be chosen from the deployed sensors or can be more powerful than other sensor nodes in terms of energy, bandwidth, and memory. Hence, the burden of transmission to the Base Station (BS) is handled by the set of cluster-heads.

- **Fault Tolerance:** Some sensor nodes may fail or blocked due to lack of power, physical damage, or environmental interference. The failure of sensor nodes should not affect the overall task of the sensor network. If more numbers of nodes fails, MAC and routing protocols must accommodate formation of new links and routes to the data collection for base stations. This may require actively adjusting transmit powers and signalling rates on the existing links to reduce energy consumption, or rerouting packets through regions of the network where more energy is available. Therefore, multiple levels of redundancy may be needed in a fault-tolerant sensor network.

- **Coverage:** can cover a limited physical area of the environment, because of its limitations in range and accuracy. Hence, it is also an important design parameter in sensor network.
- **Quality of Service, Scalability, Connectivity:** In some applications, data delivery is periodic, data will consider for certain period otherwise become useless. As the energy gets depleted, the network may be required to reduce the quality of the results in order to reduce the energy dissipation in the nodes and lengthen the total network lifetime. Energy-aware routing protocols are required to capture these requirements. The sensing area may be huge in the order of hundreds or thousands, or more sensor nodes. All protocols must work without considering the size of the network. The High node density in sensor networks precludes them from being completely isolated from each other. Therefore, sensor nodes are expected to be highly connected.

3.3Tree Based Protocols:
Tiny Aggregation TAG [11]: Tiny Aggregation service for ad-hoc sensor networks. It is a data-centric protocol, which is based on a tree structure data aggregation and is specifically designed for monitoring applications. This means that information is collect periodically from all the nodes. TAG consists of two main phases: (1) the distribution phase, where queries are disseminated to the sensors and (2) the collection phase, where the aggregated sensor readings are routed up the aggregation tree. The distribution phase passes by a broadcast message from the sink in order to organize the sensor nodes into a tree. This broadcast message is periodically sent by the sink to keep the tree structure updated regularly.

3.4Energy-Aware Distributed Aggregation Tree EADAT [7, 12]: Energy-Aware distributed heuristic algorithm to construct and maintain a Data Aggregation Tree in WSNs. In which the tree formation is by broadcasting a control message. This message is forwarded among all the sensor nodes until each node broadcasts the message once and the result is an aggregation tree rooted at the base station. After receiving this message for the first time, a sensor nodes set up its timer counts, when its communication channel is idle, sensor node chooses the node with higher residual power and shorter path to the sink as its parent during this process.

3.5Aggregation Efficiency-Aware Greedy Incremental Tree (AGIT) [13]: GIT (Greedy Incremental Tree) routing is a heuristic distributed algorithm to construct a Steiner tree on a hop-count basis and is based on directed diffusion (DD), which is a typical data-centric routings for sensor networks. Routing assumes perfect aggregation. Each source, tries to find the shortest hop from itself to the existing path tree or the sink, one by one. Each exploratory message in GIT routing involves an additional attribute is considered, to realize process. which denotes the additional cost (hop-count) from the source originating itself to the current node. GIT-like routing necessarily distributes exploratory messages in order to determine the aggregation point for the existing path tree. The exploratory message is distributed through the network according to the gradient of the corresponding interest. The message will arrive at nodes on the existing path tree. The AGIT routing, new scheme to suppress the excessive exploratory messages, which can construct a more efficient path tree than the original GIT routing and the opportunistic routing.

3.6.SRTSD (Spanning Routing Tree protocol based on State Diffusion) and SRTSD-DA (SRTSD with Data Aggregation)[4,14] : These protocols ensures the connectivity and to save energy in mobile WSNs. *Routing tree construction* consists of startup phase, diffusion phase and schedule creation phase. In the startup phase, all sensors set their state to be interrupted, but the state of sink is always connected. The sink node broadcasts message to interrupted nodes within the communication range of the sink. Nodes receive the location of sink node from the message. If the distance between the node and the sink is less than the reliable communication distance then the node ID and hop count are added into its connected nodes list (CNL). In *Diffusion phase,* the node traverses the CNL to find the best next hop node with smallest network cost. The diffusion phase continues until all nodes become connected. After diffusion phase, all the *connected* nodes form a tree whose root is sink node. In *Sensed Data Transmission Phase*, the connected sensors transmit sensed data to its next hop. The intermediate nodes receive data and forward the data directly. The SRTSD-DA algorithm is similar to SRTSD algorithm, with small changes of diffusion phase with respect network cost and sensed data transmission phase with respect to intermediate nodes, which can aggregate the data and transmit aggregated data to its next hop.

3.7.Power Efficient Data gathering and Aggregation Protocol (PEDAP)[6,15]: The goal of this is to maximize the lifetime of the network in terms of number of rounds, where each round corresponds to aggregation of data transmitted from different sensor nodes to the sink. PEDAP is a minimum spanning tree based protocol which improves the lifetime of the network even when the sink is inside the field. Prim's minimum spanning tree algorithm is employed to compute the routing paths with the sink as the root. The data packets are routed to the sink over the edges of the minimum spanning tree. In order to balance the load among the nodes, the residual energy of the nodes should be considered while aggregating the data. The PEDAP protocol requires global knowledge of the location of all nodes at the sink. The protocols operate in a centralized manner where the sink computes the routing information.

3.8Cluster based protocols:

Low-Energy Adaptive Clustering Hierarchy (LEACH)[16,17]:LEACH is an adaptive clustering-based protocol using randomized rotation of cluster-heads to evenly distribute the energy load among the sensor nodes in the network . The data will be collected by cluster heads from the nodes in the cluster and after processing and data aggregation forwards it to base station. The three important features of LEACH are Localized co-ordination and control for cluster setup, Randomized cluster head rotation, Local compression to reduce global data communication

LEACH is divided into five clusters, each cluster has a black circle represents the first cluster node, the rest of the white circle indicates a non cluster head node. Each cluster has a cluster head node, protocol randomly selecting cluster head node cycle, the energy of the entire network load equally distributed to each sensor node can achieve lower energy consumption, the purpose of improving network lifetime

3.9.PEGASIS: PEGASIS (Power-Efficient Gathering in Sensor Information Systems)[18]:It is considered an optimization of the LEACH algorithm. The key idea in PEGASIS is to form a chain among the sensor nodes so that each node will receive from and transmit to a close neighbor. The chain is constructed with a greedy algorithm. It Gathers the data and moves from node to node, eventually a designated node transmits to the Base Station. For a network running PEGASIS, it is required to form a chain that contains all nodes. The chain construction starts with the farthest node from the base station. By using a greedy algorithm, it chooses the second farthest node as its neighbor. Then the third farthest node is chosen as the second farthest nodes other neighbor. This process is repeated until the closest node to the base station is chosen as the other end of the chain. PEGASIS outperforms LEACH by eliminating the overhead of dynamic cluster formation, minimizing the distance non leader-nodes must transmit, limiting the number of transmissions and receives among all nodes, and using only one transmission to the BS per round.

3.10.TEEN: Threshold sensitive Energy Efficient sensor Network (TEEN)[19] is a hierarchical clustering protocol belongs to on-demand routing protocols category and mostly used for time crucial applications. In TEEN, nodes have two levels of cluster heads and also they follow hierarchical clustering design. After nodes have selected their cluster head, user needs to manually enter the attribute values, which will be broadcasted by cluster head in form of two parameters soft threshold and hard threshold.TEEN is a hybrid of hierarchical clustering and data-centric protocols designed for time-critical applications. It is a responsive protocol to sudden changes of some of the attributes observed in the WSN (e.g., temperature, pressure).The algorithm first goes through cluster formation. It enables CHs to impose a constraint on when the sensor should report their sensed data. After the clusters are formed, the CH broadcasts two thresholds to the nodes namely hard threshold (HT), This is a threshold value for the sensed attribute. It is the absolute value of the attribute beyond which, the node sensing this value must switch on its transmitter and report to its cluster head. and Soft threshold (ST) ,This is a small change in the value of the sensed attribute which triggers the node to switch on its transmitter and transmit. It is useful for the applications where the users can control a trade-off between energy efficiency, data accuracy, and response time dynamically.

3.11.Adaptive Periodic Threshold Sensitive Energy Efficient Sensor Network Protocol (APTEEN)[20] :It has been proposed just as an improvement to TEEN in order to overcome its limitations and shortcomings. It mainly focuses on the capturing periodic data collections (LEACH) as well as reacting to time-critical events (TEEN). Thus, APTEEN is a hybrid clustering-based routing protocol that allows the sensor to send their sensed data periodically and react to any sudden change in the value of the sensed attribute by reporting the corresponding values to their CHs. The architecture of APTEEN is same as in TEEN, which uses the concept hierarchical clustering for energy efficient communication between source sensors and the sink. APTEEN guarantees lower energy dissipation and a helps in ensuring a larger number of sensors alive. In this, cluster is formed with 1st level and 2nd level cluster heads. After selecting the cluster head (CH) it receives the attribute from the user. The CH broadcasts the attribute, hard threshold (HT), soft threshold (ST), schedule and count time (CT) parameters to the cluster members. The sensor nodes sense continuously. If the sensed value is above the HT it is stored in the internal variable (SV) and transmitted to the CH through the TDMA schedule assigned for it. All the clusters formed here may not have uniform number of sensor nodes. The cluster formed with maximum number of nodes requires more time to aggregate the captured data from nodes and transmit to BS than the cluster with minimum number of nodes. APTEEN Guarantees lower energy dissipation, It ensures that a larger number of sensors are alive.

3.12.(Hybrid Energy-Efficient Distributed Clustering(HEED)[21]:HEED is a multi-hop clustering algorithm for wireless sensor networks, with a focus on efficient clustering by proper selection of cluster heads based on the physical distance between nodes. It uses using residual energy as primary parameter and network topology features (e.g. node degree, distances to neighbors) are only used as secondary parameters to break tie between candidate cluster heads, as a metric for cluster selection to achieve load balancing. In this all nodes are assumed to be homogenous i.e. all sensor nodes are equipped with same initial energy. The node population is equipped with more energy than the rest of the nodes in the same network , this is the case of heterogeneous sensor networks. As the lifetime of sensor networks is limited there is a need to re-energize the sensor network by adding more nodes.The main objectives of HEED are to Distribute energy consumption to prolong network lifetime, Minimize energy during the cluster head selection phase, Minimize the control overhead of the network. Cluster heads are determined based on two important parameters; The residual energy of each node is used to probabilistically choose the initial set of cluster heads. This parameter is commonly used in many other clustering schemes. Intra-Cluster Communication Cost is used by nodes to determine the cluster to join. This is especially useful if a given node falls within the range of more than one cluster head.

Hierarchal Protocols	Type of Network	Energy Consumption	Aggregation	Scalability
PEDAP	Ad-hoc Network	Low	Yes	Limited
TAG	Ad-hoc Network	Low	Yes	Good
AGIT	Ad-hoc Network	Low	Yes	Limited
SRTSD,SRSTD-DA	Mobile sensor network	High	Yes	Good
EADAT	Ad-hoc Network	Low	Yes	Limited
LEACH	Ad-hoc Network	High	Yes	Good
TEEN	Ad-hoc Network	High	Yes	Good
APTEEN	Ad-hoc Network	High	Yes	Good
PEGASIS	Ad-hoc Network	High	No	Good
HEED	Ad-hoc Network	Moderate	Yes	Limited

Fig 4 Comparison of routing protocols

IV. CONCLUSION:

Data aggregation is one of the key techniques to solve the resource-constrained problem in WSN. WSN routing protocols with purpose to find the path to save energy and establish reliable data transfer method from source to destination. This paper presents an overview of different aggregation mechanisms, in tree based and cluster based routing protocols for wireless sensor networks when energy consumption and scalability is considered. There are still many problems need to be considered how to construct the aggregation tree for wireless sensor networks to maximize the life time. Survey of existing protocols was discussed in cluster based and tree based routing protocols in wireless sensor network, all of them need low processing and memory for routing that means lower energy requirements. However the selection of the protocols entirely depends upon the requirements of the application.

REFERENCES:

[1] Wen-Wen Huang, Ya-Li Peng,Jian Wen, Min Yu, “Energy-Efficient Multi-hop Hierarchical Routing Protocol for Wireless Sensor Networks”, 2009 International Conference on Networks Security, Wireless Communications and Trusted Computing.

[2] Abid Ali Minhas, Fazl-e-Hadi, Danish Sattar, KashifMustaq and S. Ali Rizvi,“ Energy efficient Multicast Routing Protocols for Wireless Sensor Networks”.

[3] Jamal N. Al-Karaki, Ahmed E. Kamal,“ Routing Techniques in Wireless Sensor networks: A Survey”.

[4] V. Akila1, Dr. T. Sheela,“ Overview of Data Aggregation Based Routing Protocols in Wireless Sensor Networks”, 2013 International Journal of Emerging Technology and Advanced Engineering.

[5] Bhaskar Krishnamachari, Deborah Estrin, Stephen Wicker,“The Impact of Data Aggregation in Wireless Sensor Networks”.

[6] Ramesh Rajagopalan, Pramod K. Varshney,“Data aggregation techniques in sensor networks: A Surevy” 2006,Electrical Engineering and Computer Science.

[7] Ahmad Sardouk1, Leila Merghem-Boulahia,“Data Aggregation in WSNS:A survey” 2010 Nova publishers.

[8] Xuxun Liu,“ A Survey on Clustering Routing Protocols in Wireless Sensor Networks”, ISSN 1424-8220, www.mdpi.com/journals/sensors.

[9] Xin Li, ShuBo Qiu,“ Research on Multicast Routing Protocol in Wireless Sensor Network”2011 IEEE.

[10] Jamal N. Al-Karaki, Ahmed E. Kamal,“ Routing Techniques in Wireless Sensor Networks: A Survey”

[11] Samuel Madden, Michael J. Franklin, and Joseph M. Hellerstein,“ TAG: a Tiny AGgregation Service for Ad-Hoc Sensor Networks”, 5th Annual Symposium on Operating Systems Design and Implementation (OSDI). December, 2002.

[12] Min Ding, Xiuzhen Cheng, Guoliang Xue “Aggregation Tree Construction in Sensor Networks” 2003 IEEE.

[13] Takafumi Aonishi, Takashi Matsuda, Shinji Mikami, Hiroshi Kawaguchi, Chikara Ohta, Masahiko Yoshimoto,“ Impact of Aggregation Efficiency on GIT Routing for Wireless Sensor Networks” 2006 International Conference on Parallel Processing Workshops.

[14] Wei Xing, Kuntai Li, Yongchao Wang, Wen Zhou,"A Spanning Routing Tree Protocol Based on State Diffusion for Mobile Sensor Networks",2011 IEEE.
[15] Huseyin O zgur Tan and I˙brahim Korpeoglu," Power Efficient Data Gathering and Aggregation in Wireless Sensor Networks".
[16] Chunjuan Wei, Junjie Yang, Yanjie Gao, Zhimei Zhang," Cluster-based Routing Protocols in Wireless Sensor Networks: A Survey", 2011 International Conference on Computer Science and Network Technology.
[17] Wendi Rabiner Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan," Energy-Efficient Communication Protocol forWireless Microsensor Networks", Proceedings of the 33rd Hawaii International Conference on System Sciences – 2000.
[18] Stephanie Lmdsey and Cauligi **S.** Raghavendra," PEGASIS: Power-Efficient GAthering in Sensor Information Systems".
[19] Arati Manjeshwar and Dharma P. Agrawal," TEEN: A Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks".
[20] Arati Manjeshwar and Dharma P. Agrawal," APTEEN: A Hybrid Protocol for Efficient Routing and Comprehensive Information Retrieval in Wireless Sensor Networks".
[21] Harneet Kour, Ajay K. Sharma,"Hybrid Energy Efficient Distributed Protocol for Heterogeneous Wireless Sensor Network",2010International Journal of Computer Applications.