A hybrid data aggregation scheme for provisioning Quality of Service (QoS) in Internet of Things (IoT)

H. Rahman[1], N. Ahmed[1], and Md. I. Hussain[2]

Department of Information Technology
North-Eastern Hill University
Shillong, India
[2]*ihussain@nehu.ac.in*
[1]{*hafizjec, nurzaman713*}*@gmail.com*

***Abstract*—Internet of Things (IoT) is a new paradigm which is enormously gaining ground in today's world. In order to maintain desirable service quality in the transmission of sensed data, data aggregation schemes are highly used. The main goal of data aggregation scheme is to collect and aggregate data packets in an efficient manner so as to reduce power consumption, traffic congestion, and to increase network lifetime, data accuracy, etc. In this paper, a hybrid Quality of service-Aware Data Aggregation (QADA) scheme is proposed. This scheme combines the features of the cluster and tree-based data aggregation schemes and addresses some of their important limitations. Simulation results show that QADA outperforms cluster and tree-based aggregation schemes in terms of power consumption, network lifetime and bearing higher traffic load.**

***Index Terms*—Data Aggregation, Internet of Things, Quality of Service, Wireless Sensor Network**

I. Introduction

Internet of Things (IoT) is the convergence of Internet, Sensors, RFID and other smart objects. Development of various applications with farsighted vision, IoT is foreseen as an integral part of the future Internet. Things in IoT provide access to real-world information. It allows interconnection of people and things anytime, anywhere, with anything or anyone using any path and any service. IoT connects trillions of smart devices seamlessly covering a variety of applications, protocols and domains [1]. Such smart devices are uniquely identifiable and addressable.

Wireless Sensor Network (WSN) is an essential part of IoT which helps in collecting information from the surroundings. It has several applications in many areas such as health monitoring, industrial automation, environment, agriculture and building automation and military applications. WSN uses a large number of tiny low-power, low-cost, having low processing capability, low memory, and multi-functional sensor nodes which are randomly and highly distributed in the physical environment [2]. Due to the deployment of extremely large numbers of devices, a huge amount of relevant, correlated and redundant data are needed to be sent by the sensors to the sink node [3]. Data generated from neighboring sensor nodes are often correlated and highly redundant. These redundant data consumes network resources unnecessarily. To overcome this kind of imprudent data transmissions in such resource-constraint network, a scheme for combining all the redundant and correlated data into valid high-quality information is needed at the intermediate nodes. This process can reduce the number of packets transmitted to the sink node ([3] and [4]). In such situations, data aggregation scheme is a suitable solution.

Various types of aggregation techniques are available in the literature. In cluster-based data aggregation, all the nodes of a cluster forward the sensed data to the Cluster Head (CH) node for aggregation. CHs aggregate data and directly forward them to the sink node for further processing. Here, energy consumption of the network increases along with the increasing distance between CHs and sink node. On the other hand, tree-based approaches reduce the distance between aggregator nodes (CH in the case of cluster-based) and sink node by constructing a logical tree among them thereby consuming lesser power than cluster-based ones. In this case, the responsibilities of aggregator node are not evenly distributed among the nodes which lessens network lifetime. To this end, we propose a hybrid data aggregation scheme which reduces network power consumption and increases network lifetime. The performance of the proposed scheme is compared with relevant ones such as LEACH, LEACH-C, and TREEPSI. The simulation results show that QADA significantly improves power consumption and network lifetime over the other protocols.

The rest of the paper is organized into following sections. Section II discusses some related works on data aggregation. Details of the proposed scheme is discussed in Section III. Section IV presents the evaluated performance of the proposed scheme vis-a-vis other relevant protocols. Finally, Section V concludes the paper.

II. Background

Data aggregation is a process of integrating and summarizing data received from sensor nodes. This process improves the network lifetime by eliminating redundant transmissions of data [5].

A plethora of cluster-based energy-efficient data aggregation schemes for WSNs is available in the literature. LEACH [6] addresses the issue of energy consumption through the distribution of CHs responsibility among all the nodes by selecting each one at different rounds. An important limitation of this scheme is that a node with little energy may be chosen as a CH which would further shorten the network lifetime. It results in unbalanced load as it does not consider nodes residual energy while selecting them as CH. To this end, LEACH-C [7] proposes a scheme with more uniformly distributed clusters than LEACH [8]. In both LEACH and LEACH-C, the CHs directly communicate with the sink node which attributes to more power consumption. A two-tier cluster-based aggregation protocol TTDCA [9] also works in a similar fashion where CHs forward the aggregated data directly to the sink node.

The tree-based architecture is more suitable for achieving energy efficiency [10]. Based on tree-based architecture, TREEPSI [11] constructs a tree considering sink as the root node. In TREEPSI, all the leaf nodes forward the data to their parents and then is rooted towards the sink. When a packet is lost at a given level of the tree, the data coming from the related sub tree are lost as well. Bahi *et al.* [12] propose tree-based aggregation protocols that work in two phases. The first phase is at local level aggregation and the second phase at the aggregators level. At each period *p*, every node forwards their aggregated data set to their proper aggregator which subsequently aggregates all data sets coming from various sensor nodes and forwards them to the sink node. Here, high load of aggregator node reduces network lifetime.

Designing a data aggregation scheme by integrating the best of both the cluster and tree-based approaches could be an interesting and efficient method to maintain the quality of data transmission.

III. The Proposed Protocol

The proposed hybrid scheme is a combination of both cluster and tree-based data aggregation techniques. It processes the raw data into a valid high-quality information, significantly reduces the redundancy of transmitting data and hence increases energy efficiency and network lifetime.

A. System Model

A WSN with a set of sensor nodes $S = \{S_1, S_2, S_3, ...S_n\}$, where n is the total number of sensor nodes, which send sensed data to the $sink$ within the network. We assume that all the sensor nodes within the network are static in nature and are randomly deployed. Every node sends their residual energy and location information to the $sink$ node. Based on this, $sink$ calculates average energy (E_avg) using Eq. (1),

$$E\_avg = \frac{(E_{s_1} + E_{s_2} + ... + E_{s_n})}{n} \tag{1}$$

If E_{s_1} is the current residual energy for a node S_1, then $sink$ selects S_1 as CH only when $E_{s_1} \geq E\_avg$ and is about nearest to the sink node. The least distance between the two sets of nodes can be calculated using the distance formula- $d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$, where d is the distance between two sets of nodes, (x_1, y_1) and (x_2, y_2) are the coordinates of the $node1$ and $node2$ respectively.

$Sink$ broadcasts a message ADV_msg to all the sensor nodes in the network. This control packet contains the header and IDentification (ID) of selected CHs (CH_ID). After receiving ADV_msg, each node compare it's ID ($nodeID$), with ADV_msg. The nodes for which ($nodeID$) = (CH_ID), become the CH nodes for that round. The total number of round T_round can be set using the Eq. (2)

$$T\_round = \frac{n}{n\_cluster} \tag{2}$$

where $n\_cluster$ is the expected number of clusters for the current round. When the current round is greater than or equal to the T_round, then it comes back to the initial round. This process repeated in each round.

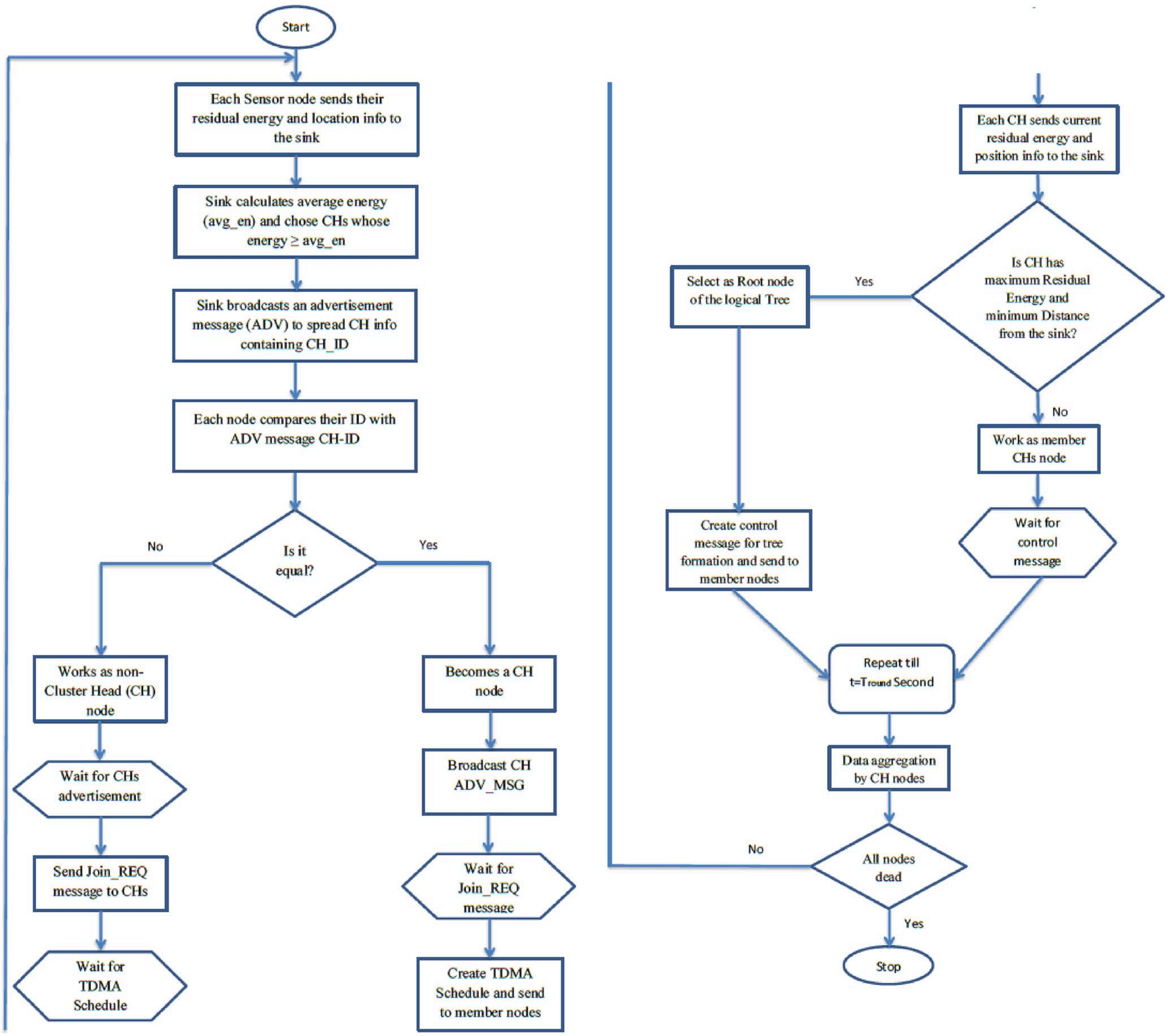
After the selection of CH by the $sink$, cluster formation is done. Each CH node broadcasts a control message CH_ADV including its ID. CH nodes also maintain a CDMA code to minimize the interference at the time of data transmission between the CHs. Based on the Received Signal Strength (RSS), non-sensor nodes send a $Join\_REQ$ message ($Join\_REQ = nodeID + CH\_ID$) to choose CH node within the cluster. CH nodes create a TDMA schedule for each non-CH node within its cluster and broadcast it to the member nodes. After cluster formation and CH selection, a logical tree among CH nodes is formed. The tree formation is initiated by the $sink$, which broadcasts the control message based on the location and current residual power of CH nodes. Data transmission starts with the non-CH nodes towards CH nodes and CHs aggregate them as and when requested. Finally, each CH node forwards the aggregated data towards the upper level CH nodes rooted at the $sink$.

B. Protocol Phases

The proposed data aggregation scheme works in four phases- (i) cluster head selection (ii) cluster formation (iii) tree formation, and (iv) data aggregation and transmission.

1) Cluster Head (CH) Selection: CH selection process is coordinated by the sink node. Initially, each node sends their residual energy and the location information to the sink node. The sink node calculates the average residual energy and selects a node as CH whose residual energy is greater than or equal to the average residual energy and approximately has more neighbor nodes than others. Sink node broadcasts an advertisement message (ADV) using CSMA MAC protocol to spread the CH information. This message contains CH node ID and a field to identify this as an announcement message. After receiving this ADV message, each sensor node compares their ID with the received one. The node which has the same ID will act as CH node for that round.

2) Cluster Formation: For cluster formation, each CH node broadcasts a message (CH-ADV) to non-cluster head nodes. Non-cluster head nodes send a join request message (Join-REQ) to the selected CH node using CSMA protocol based on the Received Signal Strength (RSS). This message includes the node's ID and the CH ID. The CH node creates a TDMA schedule and broadcasts this to the member nodes within its



(a) CH selection and cluster formation (b) Tree formation and data aggregation

Fig. 1: Flow diagram of the proposed scheme

cluster. Based on this TDMA schedule, the radio components of each non-clustered head node to be turned off excluding during their transmission time. Once, the TDMA schedule is known by all nodes in the cluster, the cluster formation phase is completed.

3) Tree Formation: After CH selection and cluster formation, the formation of the logical tree is initiated by the sink node itself. This tree construction is mainly based on the position and the current residual energy of CH nodes. At first, the sink node broadcasts a control message among all the CH nodes. This control message has five fields- ID, parent, power, status, and level, indicating the CH's ID, it's parent in the aggregation tree, it's current residual power, it's status (leaf node, relay node, or danger state) in the logical tree, and the path length (number of hops from the sink) respectively. The contents of the control message for the sink node is $msg(ID_{sink_i}, -, \infty, status_{sink_i}, level_0)$, assuming it has infinite power supply and it is the root node of the aggregation tree. The CH_1 records the parent node with higher residual power and the smallest path to the $sink$. Now CH_1 broadcasts the message $msg(ID_{CH_1}, parent_{CH_1}, power_{CH_1}, status_{CH_1}, level_{CH_1})$, where $level_{CH_1} = 1 + level_0$. This process continues until each CH broadcasts control message once. This result is an aggregation tree of CH nodes with sink at the root node. Depending on the residual power of the CH nodes, sink node can re-constructed the aggregation tree periodically. This information also contains the TDMA schedule for each CH node. The CH nodes which do not have data to send keep their radio's off to reduce power consumption. Fig. 2 shows the hybrid architecture of our proposed scheme.

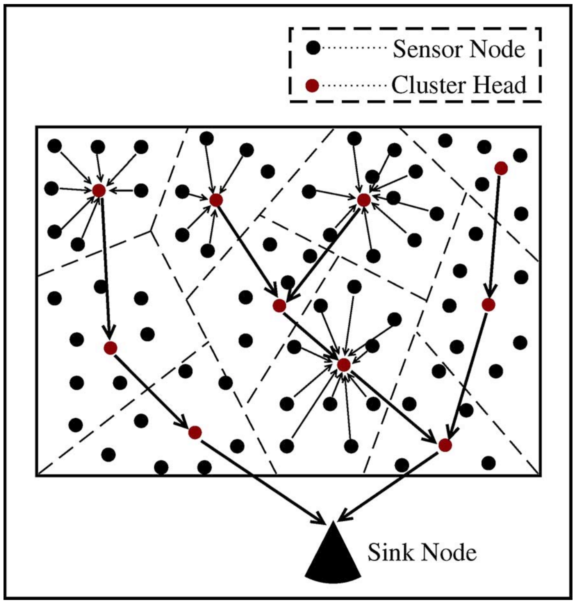


Fig. 2: An architecture of the proposed hybrid data aggregation scheme

4) Data Aggregation and Transmission: After formation of the tree among CHs, data transmission process starts. Each sensor node (non-CH) sends data packets to their respective CH nodes once per frame time. The cluster heads aggregate the data received from other nodes along with the data sensed by itself and forward them to the next higher level (parent) nodes. This process is repeated until aggregated data reaches the sink node.

Fig. 1 gives the detail flow of the proposed scheme. The CH selection and cluster formation procedures are described in Fig. 1a. Further, Fig. 1b gives the tree formation and data aggregation among the CH nodes.

IV. Performance Evaluation

Performance evaluation of QADA is carried out using NS-2.34 [13] which is presented in this section.

A. Experimental Setup

For experimental setup, we used a network of 101 nodes which randomly distributed. It is assumed that the initial energy for all the nodes is same and they have their location information. Further, all the sensor nodes are considered as static which is densely placed so that they can communicate with other nodes. The details of the parameters used in the simulation study are listed in Table I.

TABLE I: Parameters used in simulation study

Parameters	**Value**
Number of nodes	101
Simulation area	100*100 meters
Simulation time	600s
Channel bandwidth	1 Mbps
Phy	Phy/WirelessPhy
MAC	Mac/Sensor
Routing protocol	AODV
Processing delay	25 μs
Packet size	500 Bytes
Duration of each round	20 seconds
Radio propagation model	Two ray ground
Antenna model	Antenna/Omni antenna

B. Experiments Conducted

We simulate our proposed scheme and compare it's performances with some cluster and tree-based protocols (LEACH, LEACH-C and TREEPSI) through the following experiments.

1) Total energy dissipation over time: It can be observed from the Fig. 3 that total energy dissipation increases with increasing time in LEACH, LEACH-C and TREEPSI protocols. LEACH dissipates more energy than LEACH-C because of its probabilistic cluster formation and CH selection. TREEPSI consumes less energy than LEACH and LEACH-C by minimizing the distance between the nodes and the sink. QADA consumes the least energy over the cluster and tree-based schemes by taking the advantages of both schemes till the end and hence show better performance.

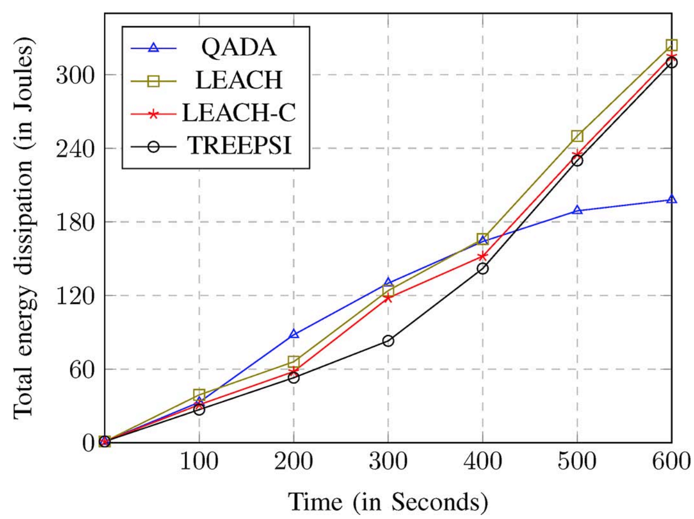


Fig. 3: Total energy dissipation over time

2) The total number of nodes alive over time: It can be seen from the Fig. 4 that with increasing time, the number of alive nodes decreases slowly in LEACH-C compared to other protocols. Because in LEACH-C, the formation of cluster and selection of CH is done by the sink node evenly. Again, QADA performs better than both LEACH and TREEPSI protocols. Due to probabilistic CH selection, LEACH's number of alive node drastically decreases over time. The rate of decrease of dead nodes in QADA is less as compared to LEACH protocol

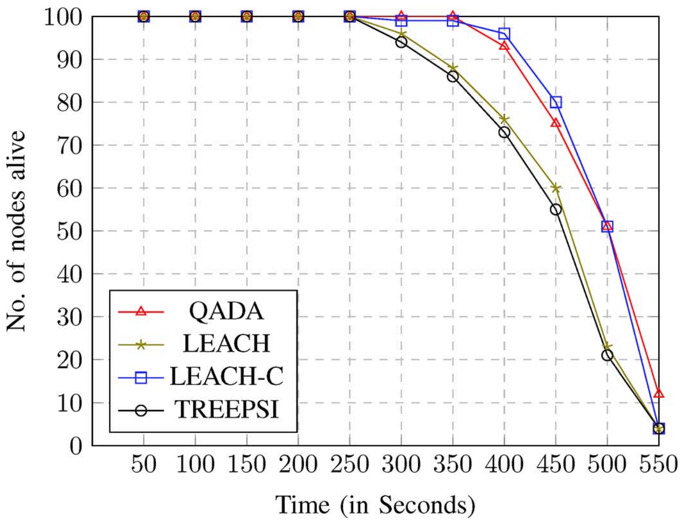


Fig. 4: Total number of nodes alive over time

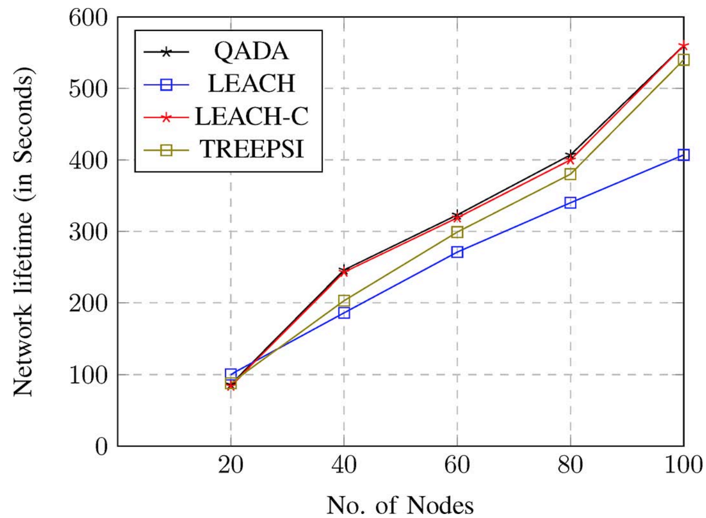


Fig. 5: Variation of network lifetime over number of nodes

which substantiates that QADA gives better performance than that of the cluster and tree-based protocols with respect to network lifetime.

3) Variation of network lifetime over the number of nodes: QADA increases network lifetime with the increasing number of nodes. Fig. 5 shows that network lifetime of cluster-based, tree-based and our proposed schemes increases along with the number of nodes, but LEACH-C obtains higher network lifetime than LEACH and TREEPSI. As in LEACH-C, cluster formation, and CHs selection is done centrally by the sink node itself. In TREEPSI, as the number of nodes increases sink node can manage the tree structure. But LEACH selects CHs and forms cluster based on probabilistic values which lead to less network lifetime with increased number of nodes. The QADA improves network lifetime over cluster and tree-based protocols as it integrates both the approaches. Hence, the dying of nodes in QADA decreases gradually.

V. Conclusion

In this paper, we have proposed a hybrid data aggregation scheme which combines some positive features of both the cluster and tree-based approaches. It drastically reduces the volume of sensed data transmitted from the end devices to the sink node. The cluster-based scheme maximizes the network lifetime by rotating CH nodes after each round. Tree-based schemes reduce the communication overhead by minimizing the distance between the nodes. The design of a data aggregation scheme considering the heterogeneous environment in the test beds scenario under the real condition is kept as a future work.

Acknowledgment

This work is supported by the project titled "QoS Provisioning in Internet of Things (IoT)" (Ref No. 13(7)/2015-CC&BT dated:28/09/2015) funded by DeitY(CC & BT), Govt. of India.

References

[1] G. Misra, V. Kumar, A. Agarwal, and K. Agarwal, "Internet of Things (IoT) —Technological Analysis and Survey on Vision, Concepts, Challenges, Innovation Directions, Technologies, and Applications," *American Journal of Electrical and Electronic Engineering*, vol. 4, no. 1, pp. 23–32, 2016.

[2] K. Keswani and A. Bhaskar, "Wireless sensor networks: A survey," *Futuristic Trends in Engineering, Science, Humanities, and Technology (FTESHT)*, pp. 1–7, 2016.

[3] S. Sirsikar and S. Anavatti, "Issues of Data Aggregation Methods in Wireless Sensor Network: A Survey," *Procedia Computer Science*, vol. 49, pp. 194 – 201, 2015.

[4] B. Karthikeyan, M. Velumani, R. Kumar, and S. Inabathini, "Analysis of data aggregation in wireless sensor network," in *2nd International Conference on Electronics and Communication Systems (ICECS)*, pp. 1435–1439, Feb 2015.

[5] A. Rajasekaran and V. Nagarajan, "Improved cluster head selection for energy efficient data aggregation in sensor networks," *International Journal of Applied Engineering Research*, vol. 11, no. 2, pp. 1379–1385, 2016.

[6] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *33rd annual Hawaii international conference on System sciences*, pp. 10–19, IEEE, 2000.

[7] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, 2002.

[8] V. Geetha, P. V. Kallapur, and S. Tellajeera, "Clustering in wireless sensor networks: Performance comparison of leach & leach-c protocols using ns2," *Procedia Technology*, vol. 4, pp. 163–170, 2012.

[9] D. Mantri, N. R. Prasad, R. Prasad, and S. Ohmori, "Two Tier Cluster based Data Aggregation (TTCDA) in wireless sensor network," in *IEEE International Conference on Advanced Networks and Telecommuncations Systems (ANTS)*, pp. 117–122, IEEE, 2012.

[10] S. Kaur and R. Gangwar, "A Study of Tree Based Data Aggregation Techniques for WSNs," *International Journal of Database Theory and Application*, vol. 9, no. 1, pp. 109–118, 2016.

[11] S. S. Satapathy and N. Sarma, "TREEPSI: Tree based Energy Efficient Protocol for Sensor Information," in *2006 IFIP International Conference on Wireless and Optical Communications Networks*, pp. 4–7, Mar 2006.

[12] J. M. Bahi, A. Makhoul, and M. Medlej, "A two tiers data aggregation scheme for periodic sensor networks," *Ad-hoc & sensor wireless networks*, vol. 21, no. 1-2, pp. 77–100, 2014.

[13] "Network Simulator-ns2." http://www.isi.edu/nsnam/ns.