IJCTA, 9(6), 2016, pp. 2809-2818
ISSN: 0974-5572

Tree Cluster Based Data Gathering Scheme (TCDGS) in Wireless Sensor Networks

Padmanaban K.* and R. Jagadeesh kannan**

ABSTRACT

Wireless Sensor Networks (WSNs) have an important issue in dynamically organizing the sensors into a network and routing the data from sensors to the base station. Clustering in WSNs is an effective technique for prolonging the network lifetime. The effectiveness of WSNs purely depends on the data collection scheme. Various data collection schemes such as multipath routing, tree, clustering and cluster tree are available for collecting data in WSNs. The existing data collection schemes fail to provide an assured reliable network in terms of mobility, traffic and end-to-end connection. Tree Cluster based Data Gathering Scheme (TCDGS) in WSN is proposed in this paper which constructs a tree based on Cluster Head (CH) location and Rendezvous Points (RPs). Data Gathering Node (DGN) in the tree does not participate in sensing, but it simply collects the data packet from the CH and delivers it to the base station. Simulation results show that TCDGS provides better quality of service while considering against other existing schemes.

Keywords: wireless sensor network, rendezvous point, data gathering node, base station, cluster head.

1. INTRODUCTION

Wireless Sensor Networks (WSNs) consists of a number of minute and low power sensors which use radio frequencies to perform distributed sensing tasks. WSNs have been widely applied in various applications, (e.g.) patient monitoring and disease analysis, pollution monitoring and source detection, equipment monitoring and fault prediction, sea searching, tide monitoring, fire detection and leakage of toxic chemicals, battlefield surveillance, radiations and gas detection [9, 10]. A large number of sensors are deployed in a Field of Interest (FoI) in such networks. The sensors are generally dropped in big numbers to guarantee reliability in such deployment.

The effectiveness of the WSNs lies in their sensing quality, energy consumption, mobility, flexibility, scalability, network lifetime, coverage etc. The goal of such WSNs deployed in the above environments is to deliver the sensing data from sensor nodes to the sink node in the communication network. Data collection becomes a vital aspect in determining the performance of such WSNs. Several network topology management routing protocols have been proposed effectively to perform such tasks primarily for data collection.

Topology management plays an important role in reducing various constraints such as limited energy, node failure, computational resource crisis, communication failure, long-range communication within a network, delay, traffic etc.

The topology inherently defines the routing path types as uncast or broadcast and it determines the size, type of packets and other overheads. Topology selection helps to enhance the performance, coverage, lifetime of the network and QoS of the network. A proficient topology ensures that neighbors are at smallest distance and reduces the probability of a packet being lost between sensor nodes. Another parameter that plays a

* Research Scholar, Department of Computer Science and Engineering, St. Peter's University, Chennai, India, *Email: Padmanaban.kphd@gmail.com*

** Professor, School of Computing Science & Engineering, VIT University, Chennai, India

major role in the performance of WSNs is energy consumption. Consumption of energy relates the transmission distance between the sensor nodes in the communication network.

The main contributions of the paper include the following key points:

- Clustering and Cluster Head (CH) selection by taking the parameters residual energy and the utility of the nodes to its neighbors.
- Data Gathering Node (DGN) and Rendezvous Points (RP) selection after certain collection rounds and balancing the energy consumption.
- Gathering the data from the CH to the DGN node and then to the base station using Tree Cluster based Data Gathering Scheme (TCDGS).

The rest of this paper is organized as follows: Section 2 describes the literature of data gathering algorithms. In Section 3, the proposed method, Tree Cluster based Data Gathering Scheme (TCDGS) in WSN is presented. The section 4 describes the simulation results and performance analysis. Finally the conclusion and future work are presented in section 5.

2. RELATED WORKS

The overall efficiency of the network can be resolved by the topology of WSNs. The types of logical topologies are classified as flat topology, chain based topology, cluster based topology, and tree based topology and cluster tree topology. Flat topology is an easy method to gather data from the remote location to sink, since it uses flooding, direct communication and gossiping. Each sensor node plays an equal role and forwards the data packet to the one-hop distance neighbor nodes or selected neighbor nodes in the communication network. Sensor Protocols for Information via Negotiation (SPIN) protocol resolves the issues of flooding, overlapping of sensing areas and resource blindness [1]. The algorithm starts the transmission with an advertisement message which is used to advertise a particular metadata in the network. The advertisement message id is followed by a request message which requests the specific data in the network. The data message carries the actual data. In SPIN protocol, the topological changes are localized, which means each node wishes to know only its one-hop neighbor nodes and it provides better performance than flooding. The data advertisement mechanism cannot assure the delivery of data in the network. Hence, it is not suitable for intrusion detection applications which require reliable data packets delivery over periodic intervals in the network.

Constrained Anisotropic Diffusion Routing (CADR) contains Information Driven Sensor Querying (IDSQ) and constrained anisotropic diffusion routing in order to query the sensor nodes and route the data [8]. Hence, each node evaluates cost/information type objective and that will forward the data packets based on the cost/information type gradient and application requirements which is ahead to minimize the latency and bandwidth and maximize the information gain in the network.

In chain topology, a sensor node is selected in the chain to act as a leader and then the remaining sensor nodes can communicate with each other along the chain path in the network. Excessive delay from distant nodes on the chain is the main demerit of this topology. Therefore collecting data on the longer chain obtains larger time. Chain Routing with Even Energy Consumption (CREEC) maintains the smooth energy consumption which includes the throwing schedule and chain establishment to increase the extended average lifetime [2]. The sink estimates the integer of throwing to be consigned to any sensor node on super round in throwing schedule. Sink constructs a new chain and updates on every super round in chain establishment to save energy on depleted sensor nodes in the network. Sink estimates the cumulative forwarding energy and arranges the nodes in the descending order to classify them into node levels. It constructs a chain topology by using the controlled Kruskal’s Minimum Spanning Tree (MST) algorithm which decides the shortest link to add them in the current working chain one by one until the end of the probability selection. The selected long chain is finally shortened using the link-trimming algorithm known as link swap.

Cluster based topology has been widely used in WSNs for data gathering, data dissemination, target tracking etc. Hybrid Energy Efficient Distributed (HEED) an energy efficient multi-hop clustering algorithm selects the CH based on the high average residual energy when compared to neighbor nodes and intra cluster communication cost in the network [3]. HEED provides uniform CH distribution across the network and load balancing. Multi-hop communication between CHs and Base Station (BS) supports energy conservation and scalability. However, more CH generation from tentative CH may leave several uncovered sensor nodes. These CHs may not have cluster members and that cannot fall on the coverage range of other CH. HEED broadcasts lot of control packets, which needs several iterations to form clusters. CH near to the sink may die earlier since these nodes have more routing packets overhead.

All the deployed sensor nodes can construct a logical tree in tree based topology. Tree based topology works normally with Depth First Search (DFS) or Breadth First Search (BFS) method. The entire data packet passes from leaf node to the parent nodes. Energy Efficient Data Collection Protocol for WSN Based on Tree (EEDCP-TB) is a tree based data collection mechanism which uses flooding avoidance scheme and cascading timing scheme to perform data aggregation in order to save the excess usage of the sensor nodes residual energy [4]. EEDCP-TB minimizes the sensor nodes energy and prolongs the network lifetime. Efficient Converge cast Tree (ECT) extends the network lifetime through weight balancing [5]. It uses distributed algorithm to construct an accurate weight balanced shortest path spanning tree with local information in the network. ECT significantly increases the throughput and network lifetime.

Cluster Tree Data Gathering Algorithm (CTDGA) gathers the data from sensor nodes with minimum energy consumption [6]. In CTDGA, the CH collects the data from corresponding cluster members. Then, a special CH is made responsible to collect the data from neighbor CHs and the data is forwarded to the sink by the tree. It improves the energy efficiency and the lifetime of the network. However, the effectiveness of cluster tree is based on the network parameters like scalability, data rate, coverage distance, traffic, cluster dimension, tree intensity, RSS and mobility.

Random Waypoint Model (RWM) describes the mobility management schemes for ad-hoc networks, sensor networks etc [7]. It is aimed to illustrate the mobile sensor nodes location, direction and velocity to change over a particular time interval in the communication network. Each sensor node on the WSNs travels from a starting coordinate position to a random ending coordinate position with a randomly generated constant velocity.

3. TREE CLUSTER BASED DATA GATHERING SCHEME

Existing algorithms has the limitation in the energy efficiency and network lifetime of the network. The proposed TCDGS scheme is shown in figure 1. The main objective of this algorithm is to improve the Quality of Service (QoS) of the communication network. TCDGS improves the energy efficiency and network lifetime of the network with minimum amount of delay.

The steps in TCDGV algorithm are as follows: The network consists of N homogeneous and static sensor nodes deployed randomly over an L × L area. The BS is located at one corner (diagonally) outside the communication network. All sensor nodes are initially provisioned with the same amount of energy and location aware abilities by using localization technology in the network. Each node has a limited communication range and the nodes are able to transmit within its transmission range. Each of them generates data with the same rate and sends them only once in each data collection round. It is assumed that the entire deployment area is full covered and well connected.

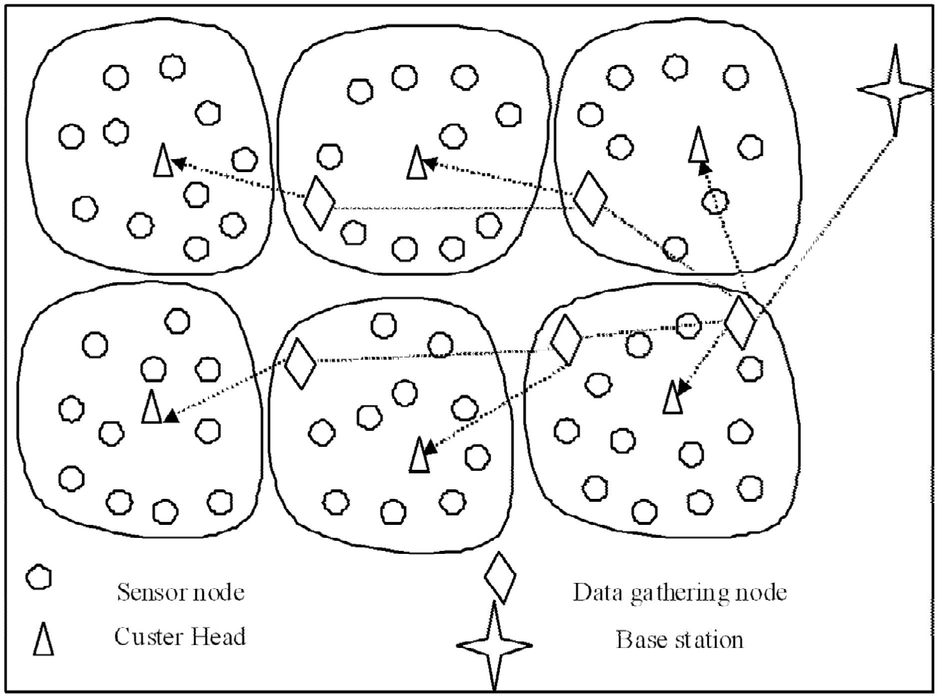

Figure 1: TCDGS scheme

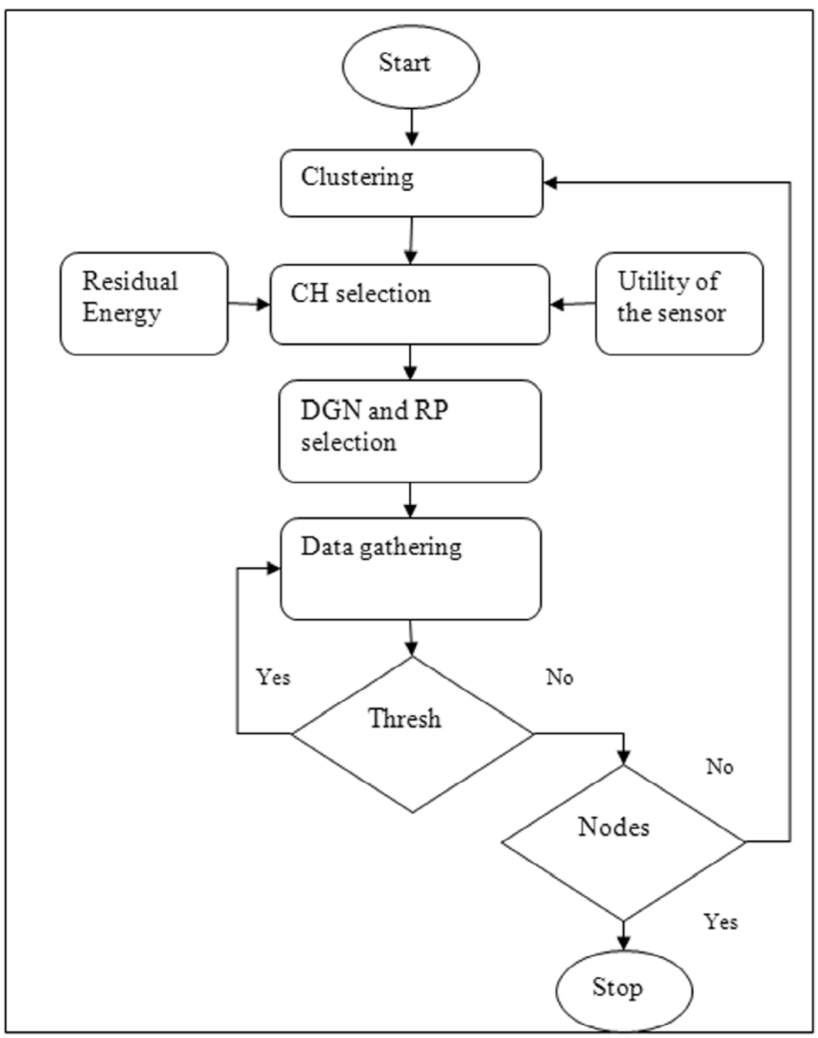

Figure 2: Flowchart of TCDGS scheme

The energy consumption of a sensor includes the power for sensing, receiving and transmission. In this paper, the energy consumption for sensing is ignored since it is negligible compared with the other two types of energy consumption. The energy consumption of one node for sending or receiving a message with *K* bits is calculated by the equations 1 and 2.

$$E_{Tx}(\mathrm{K}, d) = E_{elec} * \mathrm{K} + \varepsilon_{amp} * \mathrm{K} * \mathrm{d}^2 \tag{1}$$

$$E_{Rx}(\mathrm{K}) = E_{elec} * \mathrm{K} \tag{2}$$

Where *d* is the physical distance between two sensor nodes, and E_{elec} is the energy consumption factor representing the power per bit incurred by the transmitting or the receiving circuit and ε_{amp} represents the coefficient for the amplifier to send one bit. The flowchart description for TCDGS scheme is shown in figure 2. Initially the clustering process is started and the CH is selected based on the required parameters. The DGN and RPs are selected based on the location and the CH. The data gathering process is started from the sensor node to the CH. A threshold level is set (i.e. 0.5) to collect the data from the sensor nodes. If the threshold level is set, the data is gathered, else check about whether is node is alive or not. If the node is alive, the data gathering process is carried out so that the CH collects the data from the sensor nodes which in turn the DGN receives the data and send it to the BS.

The TCDGS algorithm consists of three stages, tree construction, RPs selection and data collection. In tree construction stage, each node is weighted by its residual energy, the number of its two-hop neighbors, its distance to the BS and the average residual energy of its one-hop neighbors. The other parameters for CH selection include coverage time, Received Signal Strength (RSS), robustness and connection time. By communicating with its one-hop neighbors and comparing their weight, every node treats the neighbor with the maximum weight as its parent node. In the initial phase, a sensor node elect itself as a CH to form a cluster, then the CH is responsible to collect the data from its cluster members. Thus, the data gathering tree can finally be constructed.

Every root node of the trees is considered as a RP. The sink initiates the Data Gathering Tree (DGT) formation process. Based on the location of CH and connection time, a few numbers of nodes are selected as DGN to generate DGT. It is represented in DGT construction algorithm. The DGN does not participate in sensing and is not a part of any cluster on that particular round and therefore it may act as ordinary sensor node in the network. The selection of DGN does not affect the collection of data from a particular cluster. The DGN collects the data from the CH and deliver it to the BS. The main goal of selecting RPs is to alleviate the "hotspot problem" so as to balance the tradeoff between the energy consumption and data gathering latency. When choosing the parameters of one node's weight, the residual energy and average energy of its one-hop neighbors should be clearly focused to avoid the fast depletion of the heavily loaded nodes in the communication network.

The sink selects the DCN based on the threshold value, connection time, RSS, communication range and heftiness for connection which reduces the surplus energy usage and traffic of the whole network. The above selected DCN can keep the communication with the CH for a longer time while the sensor nodes are on high mobility and there is no need to update in the tree structure. In order to keep the lifetime of whole network in harmony, new DCN is selected every time when the new CHs are elected. Use of threshold mechanism, also increase the number of nodes alive, means it increase the network lifetime as compared to others. It protected the parent node death slowly, because each node has chance to become a parent node. Once the data gathering process is completed, analysis can be performed by varying the location of the BS and thus the efficiency of the algorithm can be checked.

4. PERFORMANCE EVALUATION

The performance of TCDGS is analyzed by using the Network Simulator version-2 (NS2). NS2 is an open source programming language written both in C++ which is used in back end and Object oriented Tool

Command Language (OTCL) which is used in front end. NS2 is a discrete event time driven simulator which is used to mainly model the network protocols. The nodes are distributed in the simulation environment in the communication network. The parameters used for the simulation of TCDGS are shown in Table 1. The simulation of the proposed TCDGS has 50 nodes deployed in the simulation area 1000 × 1000m.

The nodes are communicated with each other by using the communication protocol User Datagram Protocol (UDP). The nodes are moved randomly within the simulation area by using the mobility model Random Way Point (RWP). The radio wave is propagated by using two ray ground propagation models. The traffic in the network is handled using the traffic model Constant Bit Rate (CBR). All the nodes receive the signal from all direction by using the Omni directional antenna. The performance of the TCDGS scheme is evaluated by using the parameters packet delivery rate, packet loss rate, average delay, throughput, residual energy and network lifetime.

4.1. Packet Delivery Rate

Packet Delivery Rate (PDR) is the ratio of number of packets delivered to all receivers to the number of data packets sent by the source node. The PDR is calculated by the equation 3.

$$PDR = \frac{Total\ Packets\ Received}{Total\ Packets\ Send} \tag{3}$$

Table 1
Simulation parameters of TCDGS

Parameter	Value
Channel Type	Wireless Channel
Simulation Time	50 s
Number of nodes	50
MAC type	802.11
Traffic model	CBR
Simulation Area	1000 × 1000
Transmission range	250m
Network interface Type	WirelessPhy
Mobility Model	Random Way Point

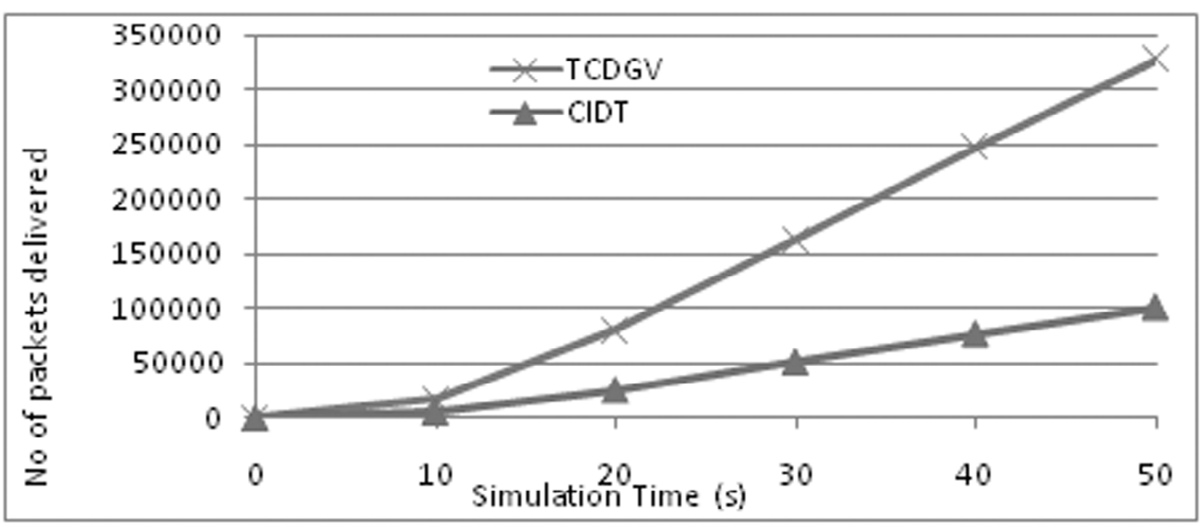

Figure 3: Packet Delivery Rate

The figure 3 shows the PDR of the proposed scheme TCDGV is higher than the PDR of the existing method CIDT. The greater value of PDR means the better performance of the protocol.

4.2. Packet Loss Rate

The Packet Loss Rate (PLR) is the ratio of the number of packets dropped to the number of data packets sent. The formula used to calculate the PLR is given in equation 4.

$$PLR = \frac{Total\ Packets\ Dropped}{Total\ Packets\ Send} \tag{4}$$

The PLR of the proposed scheme TCDGV is lower than the existing scheme CIDT in Figure 4. Lower the PLR indicates the higher performance of the network.

4.3. Average Delay

The average delay is defined as the difference between the current packets received time and the previous packet received time. The delay in the network degrades the performance of the network. The average delay is measured by equation 5.

$$Avg\_Delay = \frac{\sum_{0}^{n} Pkt\,\mathrm{Re}cvd\ Time - Pkt\ Send\ Time}{Time} \tag{5}$$

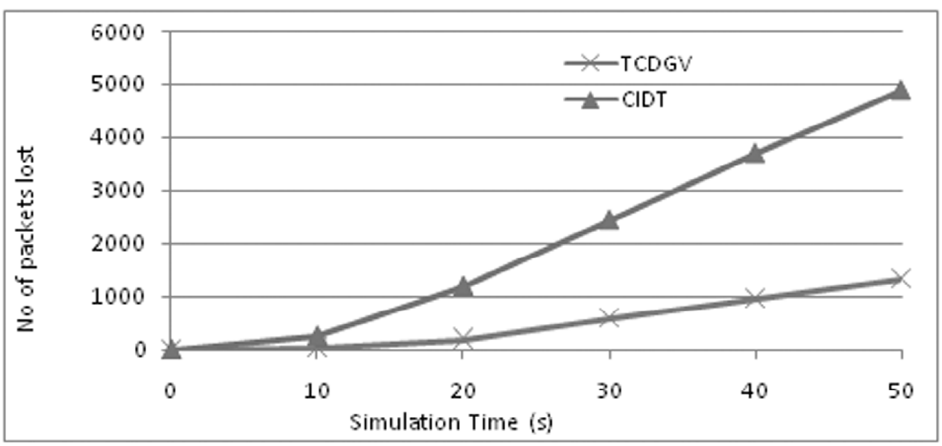

Figure 4: Packet Loss Rate

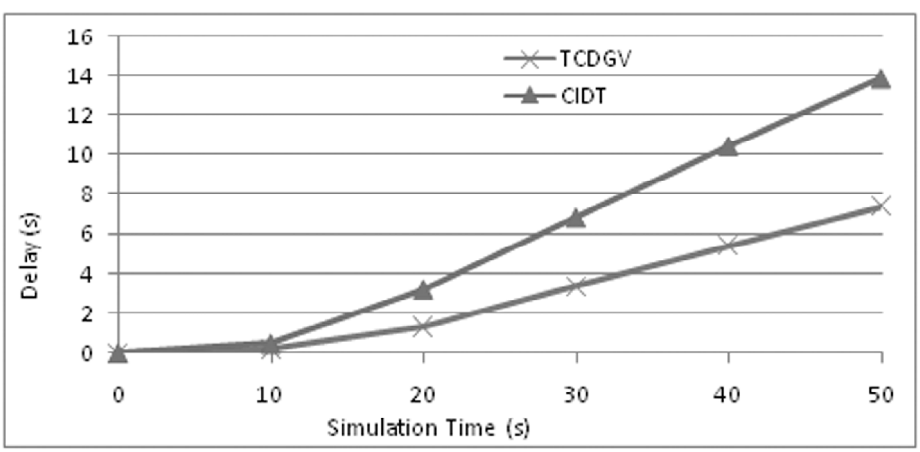

Figure 5: Average delay

Figure 5 shows that the delay value is low for the proposed scheme TCDGV than the existing scheme CIDT. The minimum value of delay means that higher value of the throughput of the network.

4.4. Throughput

Throughput is the average of successful messages delivered to the destination. The average throughput is estimated using equation 6.

$$Throughput = \frac{\sum_0^n Pkts\ Received\ (n) * Pkt\ Size}{1000} \tag{6}$$

Figure 6 shows that proposed scheme TCDGV has greater average throughput when compared to the existing scheme CIDT.

4.5. Residual Energy

The amount of energy remaining in a node at the current instance of time is called as residual energy. A measure of the residual energy gives the rate at which energy is consumed by the network operations.

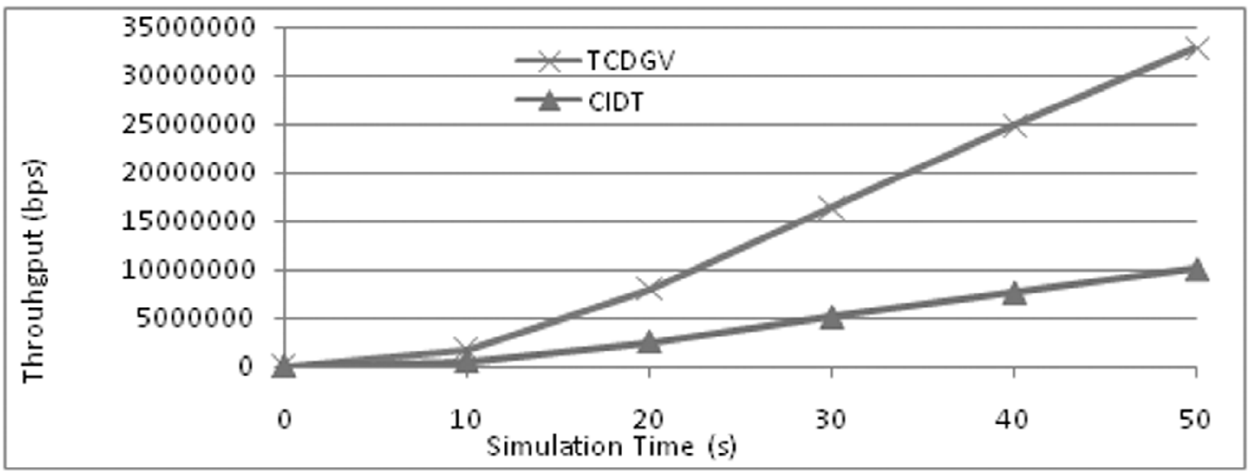

Figure 6: Throughput

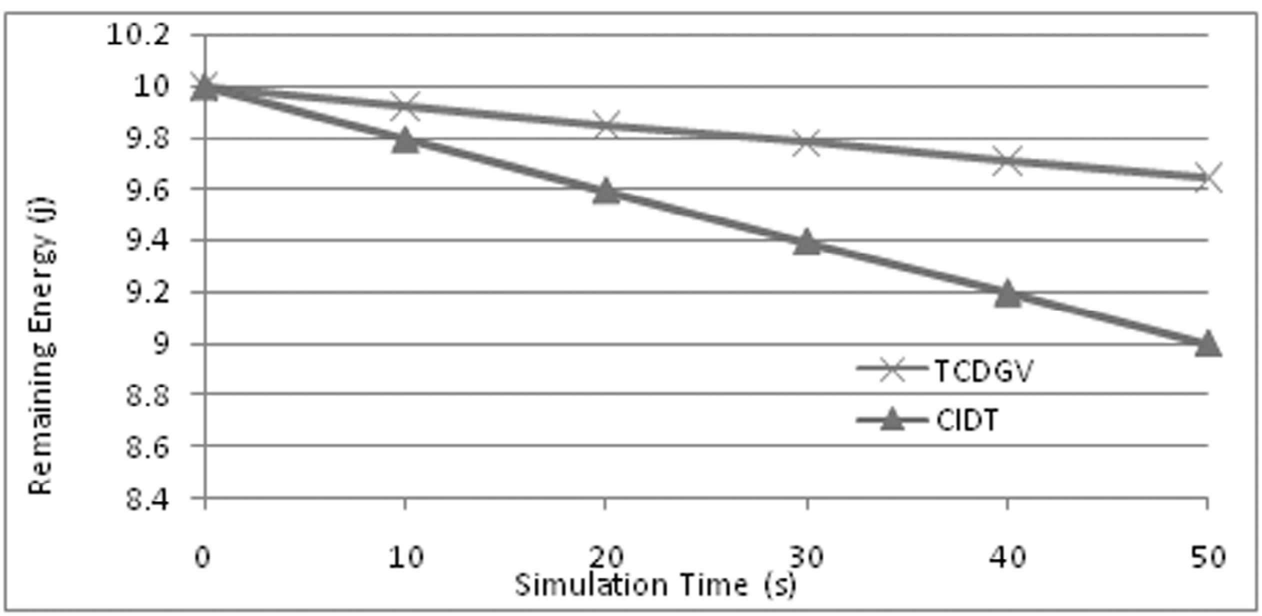

Figure 7: Residual energy

The residual energy of the communication network is better for the proposed TCDGV when compared with the existing scheme CIDT as in figure 7.

4.6. Network Lifetime

Network lifetime is defined as the number of the data collection round until the first sensor node of the network dies as a result of depleting its energy resources.

Figure 8 shows that both the TCDGV and the CIDT curves fall steeply after the death of the first node. This is a clear indication that there is maximum utilization of nodes' energy before the entire network dies.

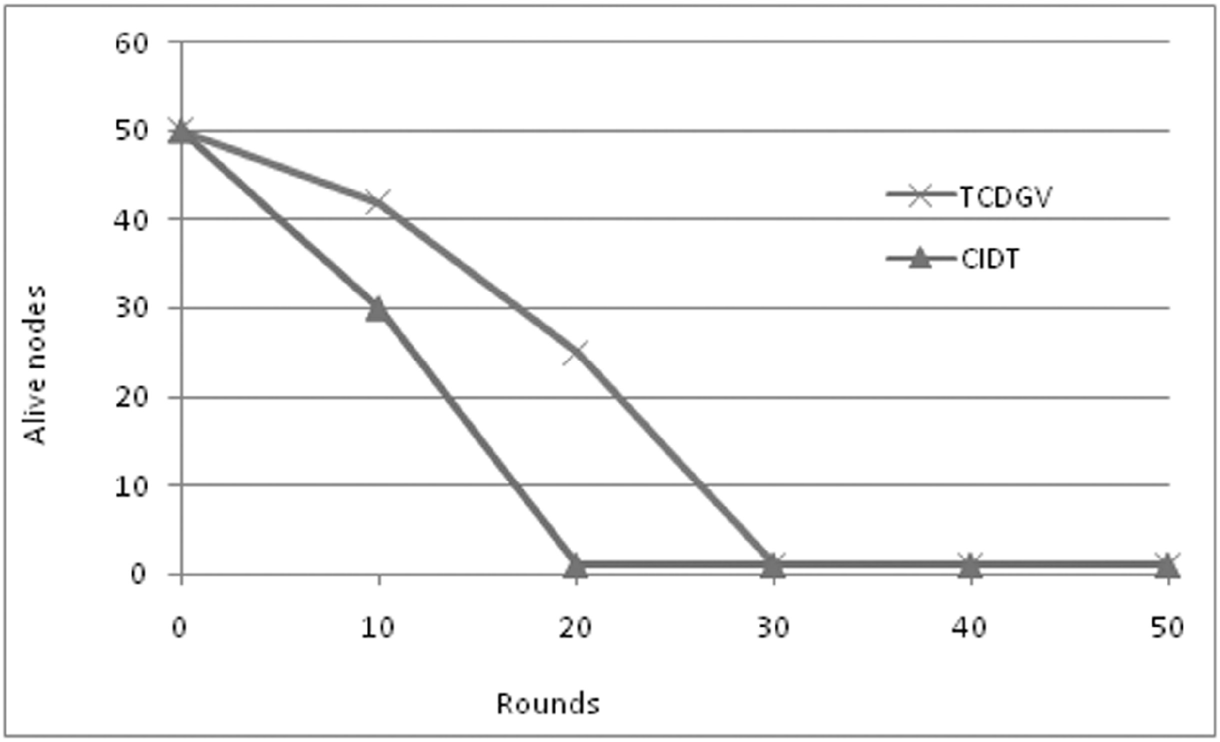

Figure 8: Network Lifetime

5. CONCLUSION

In this paper, Tree cluster based Data Gathering Scheme (TCDGS) is proposed for WSNs to exploit the network lifetime, connection time, residual energy, RSSI, throughput, PDR and stable link for sensor nodes, whereas each cluster member chooses the CH with better connection time and forwards the data packets to the corresponding CH in an allocated time slot. Similarly, the sink or DGN elects the one-hop neighbor DGN or CH with maximum threshold value, connection time, RSSI and with less network traffic. From the simulation results, it is evident that TCDGS provides more stable links, better throughput, energy utilization and PDR with reduced network traffic than CIDT.

REFERENCES

[1] W.B. Heinzelman, J. Kulik and H. Balakrishnan, "Adaptive protocols for information dissemination in wireless sensor networks," in Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom'99), Seattle, WA USA, 1999, pp. 174-185.

[2] J. Shin and C. Sun, "CREEC: Chain routing with even energy consumption," IEEE Journal of Communications and Networks, vol. 13, no. 1, pp. 17-25, 2011.

[3] 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, pp. 366-379, 2004.

[4] Y. Jin, F. Chen, G. Che, and W. Hu, "Energy Efficient Data Collection Protocol for Wireless Sensor Network Based on Tree," in Proceedings of the 2010 Asia-Pacific Conference on Wearable Computing Systems (APWCS), Shenzhen, 2010, pp. 82-85.

[5] M.R. Saadat, and G. Mirjalily, "Efficient Convergecast Tree for data collection in wireless sensor Networks," in Proceedings of the 20th Iranian Conference on Electrical Engineering (ICEE), Tehran, Iran, 2012, pp. 1534-1539.

[6] J. Yang, B. Bai, and H. Li, "A cluster-tree based data gathering algorithm for wireless sensor networks," in Proceedings of the International Conference on Automatic Control and Artificial Intelligence (ACAI2012), Xiamen, 2012, pp. 22-25.

[7] E. Hyytia, H. Koskinen, P. Lassila, A. Penttinen, and J. Virtamo, "Random Waypoint Model in Wireless Networks," 2005. Networking Laboratory, Helsinki University of Technology, Finland.

[8] M. Chu, H. Haussecker and F. Zhao, "Scalable information driven sensor querying and routing for ad hoc heterogeneous sensor networks," The International Journal of High Performance Computing Applications, vol. 16, no. 3, 293-313, 2002.

[9] T. M.Chiwewe and G. P. Hancke, "A distributed topology control technique for low interference and energy efficiency in wireless sensor networks," IEEE Trans. Ind. Informat., vol. 8, no. 1, pp. 11-19, Feb. 2012.

[10] J. S. Lee, "A Petri net design of command filters for semi-autonomous mobile sensor networks," IEEE Trans. Ind. Electron., vol. 55, no. 4, pp. 1835-1841, Apr. 2008.