

Research Article

A Hybrid Structure for Data Aggregation in Wireless Sensor Network

Hedieh Sajedi¹ and Zahra Saadati²

¹Department of Mathematics, Statistics and Computer Science, College of Science, University of Tehran, Iran

²Department of Computer Engineering, College of Technology and Engineering, Islamic Azad University, Omidiyeh Branch, Iran

Correspondence should be addressed to Hedieh Sajedi; hhsajedi@ut.ac.ir

Received 12 July 2014; Revised 7 December 2014; Accepted 9 December 2014; Published 23 December 2014

Academic Editor: Hongli Dong

Copyright © 2014 H. Sajedi and Z. Saadati. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, wireless sensor networks have been used for various applications such as environmental monitoring, military and medical applications. A wireless sensor network uses a large number of sensor nodes that continuously collect and send data from a specific region to a base station. Data from sensors are collected from the study area in the common scenario of sensor networks. Afterward, sensed data is sent to the base station. However, neighboring sensors often lead to redundancy of data. Transmission of redundant data to the base station consumes energy and produces traffic, because process is run in a large network. Data aggregation was proposed in order to reduce redundancy in data transformation and traffic. The most popular communication protocol in this field is cluster based data aggregation. Clustering causes energy balance, but sometimes energy consumption is not efficient due to the long distance between cluster heads and base station. In another communication protocol, which is based on a tree construction, because of the short distance between the sensors, energy consumption is low. In this data aggregation approach, since each sensor node is considered as one of the vertices of a tree, the depth of tree is usually high. In this paper, an efficient hierarchical hybrid approach for data aggregation is presented. It reduces energy consumption based on clustering and minimum spanning tree. The benefit of combining clustering and tree structure is reducing the disadvantages of previous structures. The proposed method firstly employs clustering algorithm and then a minimum spanning tree is constructed based on cluster heads. Our proposed method was compared to LEACH which is a well-known data aggregation method in terms of energy consumption and the amount of energy remaining in each sensor network lifetime. Simulation results indicate that our proposed method is more efficient than LEACH algorithm considering energy consumption.

1. Introduction

The manufacturing of small, low-cost, self-contained, inexpensive, and battery-powered sensors has become technically and economically feasible due to recent technological developments. Wireless sensor network (WSN) is an ad hoc network, which has wide applications in a broad range of areas including industrial, military, environment, and medical systems. A sensor node includes sensing component, data processing component, and communication unit. The sensor nodes are usually deployed in large numbers (several hundreds) and in areas where it is difficult or impossible. In most of the applications, which use the sensor nodes to monitor the remote field, there is no chance for maintenance and battery replacement. In WSN, sensor nodes are usually

scattered randomly [1]. The lifetime of each sensor node is determined by its power-source which significantly affected the relationship between the nodes. In a common sensor network scenario, data are collected by the sensor nodes from the study area and neighboring sensor nodes generate redundant data. Under this scenario, the redundant data do not need to be transferred to the base station. Thus, a method is needed for gathering and combining data that avoids redundant data transformations and in turn saves energy and bandwidth.

Data aggregation is a data transfer technique that combines data from different sensor nodes into one packet. Data aggregation combines data coming from several nodes, eliminates redundancy, minimizes the number of transmissions, and thus saves energy. The main purpose of data aggregation

is reduction of communication at different levels and in turn reduction of total energy consumption. There are different data aggregations techniques and each dissipated different amounts of energy to process raw data. The choice of data aggregation techniques depends on application requirements and energy savings of the methods. There are two well-known protocols: cluster based data aggregation and tree based data aggregation.

Clustering reduces energy dissipation and collision in local cluster but energy dissipation for cluster heads to communicate with the base station is high. In tree based method according to low distance between each of the nodes and parents, energy dissipation is low but the depth of tree is high.

In this paper, we proposed a hybrid schema called CTDA, which employs benefits of both cluster based and tree based data aggregation methods. Proposed schema is based on LEACH (low-energy adaptive clustering hierarchy) protocol, which is one of the clustering based data aggregation protocols, and minimum spanning tree between the cluster heads. This study presented a combination method, which preserves advantages of the mentioned methods and minimizes disadvantages of the clustering and tree based approaches. Specifically, CTDA is energy efficient and consists of two main phases: (1) set-up phase, which includes cluster head selection step, cluster formation step, and tree construction of cluster heads step, and (2) steady-state phase, which includes data transmission.

Our proposed method is compared to LEACH which is a well-known data aggregation method in terms of energy consumption and the amount of energy remaining in each sensor network lifetime. Simulation results indicate that our proposed method is more efficient than LEACH algorithm considering energy consumption.

The organization of this paper is as follows: in Section 2 we review the related works and Section 3 is about background. The proposed approach is described in Section 4. Section 5 explains the simulation and the results and finally Section 6 concludes the paper.

2. Related Works

In flat networks, all sensor nodes have the same role and are equipped with the same battery. In this network, data aggregation is performed by data centric routing where the base station usually transmits a query message to the sensors. In view of scalability and energy efficiency, several hierarchical data aggregation approaches have been proposed.

Hierarchical data aggregation involves data combination at particular nodes, which reduces the number of messages transmitted to the base station [2]. This improves the energy efficiency of the network. In the rest of this section, we will review two important hierarchical data aggregation protocols.

Data Aggregation in Cluster Based Networks. Cluster based data aggregation is addressed by many researches. Clustering is a fundamental solution to energy efficiency in sensor networks. Sensor nodes are organized into clusters, with each cluster having a “cluster head” as the manager.

Communication within a cluster must go through the cluster head, which is then forwarded to a neighboring cluster head until it reaches the base station [2]. LEACH [3] is one of the primary and popular clustering algorithms. The operation of LEACH is divided into rounds. Each round begins with a set-up phase when the clusters are organized, followed by a steady-state phase when data are transferred from the nodes to the cluster head and to the base station. Nodes within a cluster that have more energy are delegated as cluster heads more often than nodes with less energy.

Tree Based Data Aggregation. In tree based data aggregation (e.g., a minimum spanning tree) base stations are roots and source nodes are considered as leaves. Each node has a parent node to forward its data [4]. Flow of data starts from leaves nodes up to the base station and aggregation is done by parent nodes.

Cluster Based and Tree Based Power Efficient Data Collection and Aggregation Protocol (CTPEDCA) [5]. This method provides an energy efficient mechanism for data transmission between cluster heads. In this study, a cluster head with maximum residual energy is selected (CH_0) as the root. CH_0 creates minimum spanning tree among all cluster heads and broadcasts tree information to all cluster heads. When the network has K cluster heads, $K-1$ cluster heads (except CH_0) send data only to CH_0 and CH_0 transmits data to the base station. The disadvantage of this method is a ruin tree in which CH_0 does not have enough energy and this method is used when the base station is too far.

Cluster-Tree Based Data Gathering. In this approach [6], the base station forms the primal clusters and these clusters do not change much because all sensor nodes are immobile, whereas the selected cluster head in the same cluster may be different in each round. For a node to be a cluster head, it has to be located at the center of a cluster. Once a node is selected to be a cluster head, it broadcasts a message in the network and invites the other nodes to join its cluster. When the cluster head receives the join message from its neighbor node, it assigns the node a time slot to transmit data. When the first round is over and the primal cluster topology is formed, the base station is no longer responsible for selecting the cluster head. Base station will collect the information that cluster head had labeled in each cluster and build path in minimum spanning tree to compute the tree path.

3. Background

In clustering routing algorithms for wireless networks, LEACH is well known because it is simple and efficient. It is an adaptive clustering routing protocol proposed by Heinzelman et al. [3]. This algorithm selects cluster head nodes randomly and the rest nodes are formed based on the received signal power from cluster head nodes. The implementation procedure of LEACH includes various rounds. Each round consists of a set-up phase and a steady-state data transmission phase. In the set-up phase, the cluster head nodes are randomly selected from all the sensor nodes and several clusters are

constructed dynamically. In the steady-state data transmission phase, member nodes in every cluster send data to their own local cluster head; the cluster head compresses the data that it received from member nodes and sends the aggregated data to the base station. LEACH protocol periodically selects the cluster head nodes, renewing clustering according to round time, such that energy dissipation of each node in the network is about equal. LEACH divides the whole network into several clusters, and the run time of the network is broken into various rounds. In the LEACH protocol, all the sensor nodes have the same probability to be a cluster head, which makes the nodes in the network consume energy in a relatively balanced way to prolong network lifetime.

A spanning tree is a graph that spans all the nodes as vertices and contains no cycles. The tree is structured in the way that the node with the smallest identifier is chosen as the root. All other nodes are connected to this selected root via the shortest-path. The protocol requires each node to exchange configuration messages in a format that contains its own identifier, its selected root, and the distance to this selected root. Each node updates its configuration message upon identifying a root with a smaller identifier or the shortest-path neighbor. Furthermore, the neighbor, which the shortest-path configuration message comes from, is chosen as the parent of a node whenever it is detected.

4. The Proposed Algorithm

In this paper, we propose a hybrid cluster and tree based algorithm for data aggregation called CTDA. It reduces data transfer volume and hence enhances energy efficiency. This algorithm reduces the number of nodes, which directly send data to the base station. The basic idea of cluster based routing is to use the data aggregation mechanism in the cluster head to lessen the amount of data transmission. Therefore, CTDA reduces the energy dissipation in communication and achieves the purpose of saving energy of the sensor nodes. This scheme has four phases in each round: (1) cluster head selection, (2) cluster formation, (3) tree formation of cluster heads, and (4) data transmission.

In LEACH, the number of cluster heads in the network is selected with a priori knowledge. The number of cluster heads depends on several parameters, that is, network topology and the relative costs of calculation versus communication. However, there is no definite method to determine the optimal number of cluster heads and the relation between the number of cluster heads and whole sensor nodes will have extra impact on the network performance.

Radio Model. Recently, much research has been done about low-energy radios. There are different assumptions about the radio features, such as energy dissipation in transmitter and receiver modes, which the advantages of protocols will change. In this study, a simple model is used for the radio energy dissipation. This model includes transmitter, power amplifier, and receiver which dissipates energy to run the radio electronics [7]. In addition, both the free space (d^2 power loss) and the multipath fading (d^4 power loss) channel

models, which depend on the distance between the transmitter and receiver, are used. In general, the free space (fs) model was used when the distance is less than a threshold d_0 ; otherwise the multipath (mp) model is used [7]. Therefore, the energy expended by the radio E_{TX} when transmitting a bit of data message for a distance d can be expressed as

$$E_{TX}(l, d) = E_{TX} - \text{elec}(l) + E_{TX} - \text{amp}(l, d), \quad (1)$$

$$E_{TX}(l, d) = \begin{cases} lE_{\text{elec}} + l \cdot \varepsilon_{\text{fs}} \cdot d^2 & d < d_0 \\ lE_{\text{elec}} + l \cdot \varepsilon_{\text{mp}} \cdot d^4 & d > d_0, \end{cases} \quad (2)$$

where ε_{fs} is the energy consumed by the amplifier to transmit at a shorter distance. ε_{mp} is the energy consumed by the amplifier to transmit at a longer distance. E_{elec} indicates the energy consumed in the electronics circuit to transmit or receive the signal, which depends on factors such as the digital coding, modulation, filtering, and spreading of the signal. To receive this message, the radio expends energy E_{RX} according to (3) as in [3]:

$$E_{RX}(l) = E_{RX} - \text{elec}(l) = lE_{\text{elec}}. \quad (3)$$

For all the nodes of network, the following assumptions are considered.

- (i) All nodes have the same initial energy.
- (ii) All the nodes know their location information.
- (iii) The network has K cluster head nodes. According to the radio model, which is described, only cluster head nodes use the multipath fading channel models, while the other nodes use the free space channel models.

4.1. Set-Up Phase. In CTDA algorithm, set-up phase consists of three steps: (1) cluster head selection step, (2) cluster formation step, and (3) tree construction of cluster head step.

Step 1 (cluster head selection). The cluster head selection algorithm is the same as LEACH. In this algorithm the sensor nodes lack a centralized control from the base station which generates a random number between 0 and 1. If the random number is less than a threshold $T(n)$, the sensor nodes will broadcast an announcement message to notify other nodes that it is a cluster head [3]. In each round, if a node has been elected as a cluster head, its $T(n)$ is set to zero, so that the node will not be elected as a cluster head again. $T(n)$ can be expressed according to (4) as in [3]:

$$T(n) = \begin{cases} \frac{P}{1 - P * [r \bmod (1/P)]}, & n \in G \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

where P is the percentage of the number of clusters in the network (usually P is 0.05), r is the current round, $r \bmod (1/P)$ is the number of nodes which have been elected as cluster heads in round r , G is the set of nodes that have not been elected as cluster heads in round r , and n is the total number of nodes.

Step 2 (cluster formation). After cluster head selection, the cluster head broadcasts its identity message to noncluster head nodes in network. The cluster head node informs all other nodes in the network that it has been chosen as the head for this current round. To do so, the cluster head node broadcasts an advertisement message (ADV) using the carrier-sense multiple access (CSMA) MAC protocol. This message is a small message that includes the node's ID and a header that distinguishes this message as an announcement message. All other nodes determine their cluster for this round by choosing the cluster head that requires the minimum communication energy [3]. Assuming symmetric propagation channels for pure signal strength, the cluster head ADV with the largest signal strength is the cluster head that requires the minimum amount of transmit energy to communicate.

After each node has selected a cluster, it must notify the cluster head that it will be a member of the cluster. Each node transmits a join request message (Join-REQ) to the chosen cluster head using CSMA protocol. This message is again a short message, including the node's ID and the cluster head's ID.

The cluster heads in LEACH act as local control centers to coordinate the data transmissions in their cluster. The cluster head node sets up a TDMA (time division multiple access) schedule and transmits this schedule to the nodes in its cluster. This ensures that there are no collisions among data messages and allows the radio components of each noncluster head node to be turned off at all times except during their transmit time, thus reducing the energy consumed by the individual sensors. Once, the TDMA schedule is known by all nodes in the cluster, the set-up phase is completed and the steady-state phase (data transmission) can start.

Step 3 (tree construction of cluster heads). After cluster formation, the cluster head broadcasts its ADV, including the cluster head ID, location, cluster ID, cluster size (i.e., the number of nodes in local cluster), and residual energy. Then, each cluster head also sends its data and location to the base station. The base station based on the position of cluster head nodes provides a minimum spanning tree between them and finally broadcast tree information for all cluster head nodes that it also inclusive a schedule for data transmission among the cluster heads. In this scheme, a minimum spanning tree is created among all the cluster heads to achieve the goal, which cluster heads used free space channel model to send data to base station. In this tree, in each round, based on location of cluster head nodes choose the minimum distance from a vertex in the tree to another vertex. An example of tree construction from cluster heads is shown in Figure 1.

4.2. Steady-State Phase (Data Transmission). Data aggregation combines data from multiple sensors to eliminate redundant transmission. Data transmission is broken into frames where noncluster head nodes send their data to the cluster head once per frame during their allocated transmission slot. Thus, nodes transmit their data without collisions within the network. In this research, we assumed that nodes are all

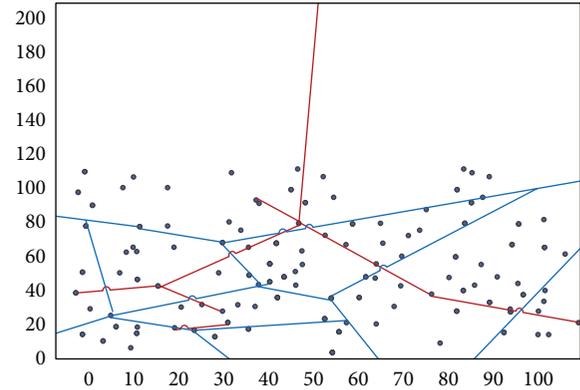


FIGURE 1: Tree construction of cluster heads.

time synchronized by having the base station sending out synchronization pulses to each node.

Each noncluster head node uses power control to send data to reduce energy dissipation. To further conserve energy, the radio of each noncluster head node is turned off until its allocated transmission time. At the allocated time, the cluster head is awakened to receive all the data from the nodes in the local cluster. Once the cluster head receives all the data from local noncluster head nodes, it performs data aggregation to generate beneficial data messages. After all the cluster heads complete the data aggregation of the local cluster, they will transmit their resultant data along the tree (by the minimum spanning tree between cluster head nodes) and each receiver node will send its result data after execution of data aggregation for local resultant data and data received from sender node. The base station from the minimum spanning tree root receives the final resultant data. The flowchart of steady-state phase in this proposed algorithm is shown in Figure 2.

5. Simulation and Results

We simulated and evaluated the performance of the proposed protocol. We used three different routing protocols: LEACH protocol, our proposed method CTDA algorithm, and MCTDA (modified clustering and tree based data aggregation) algorithm. In MCTDA, the minimum spanning tree does not aggregate data and instead of it cluster heads send data to the base station. We used the same simulation environment as LEACH, to facilitate comparison of energy savings between the tested protocols. Hundred sensor nodes are randomly deployed in a field with dimension $100\text{ m} \times 100\text{ m}$. Dots in Figure 3 show these nodes. The base station is located far away from the region, at coordinates (50, 200) which is shown with black diamond in Figure 3. The parameters and the values used to develop the simulation ((1) and (3)) are listed in Table 1.

In this simulation, each node has 0.5 J of energy at the beginning of the experiment and a limitless amount of data to send to the base station. When a node uses up its allocated energy during the experiment, it no longer transmits or

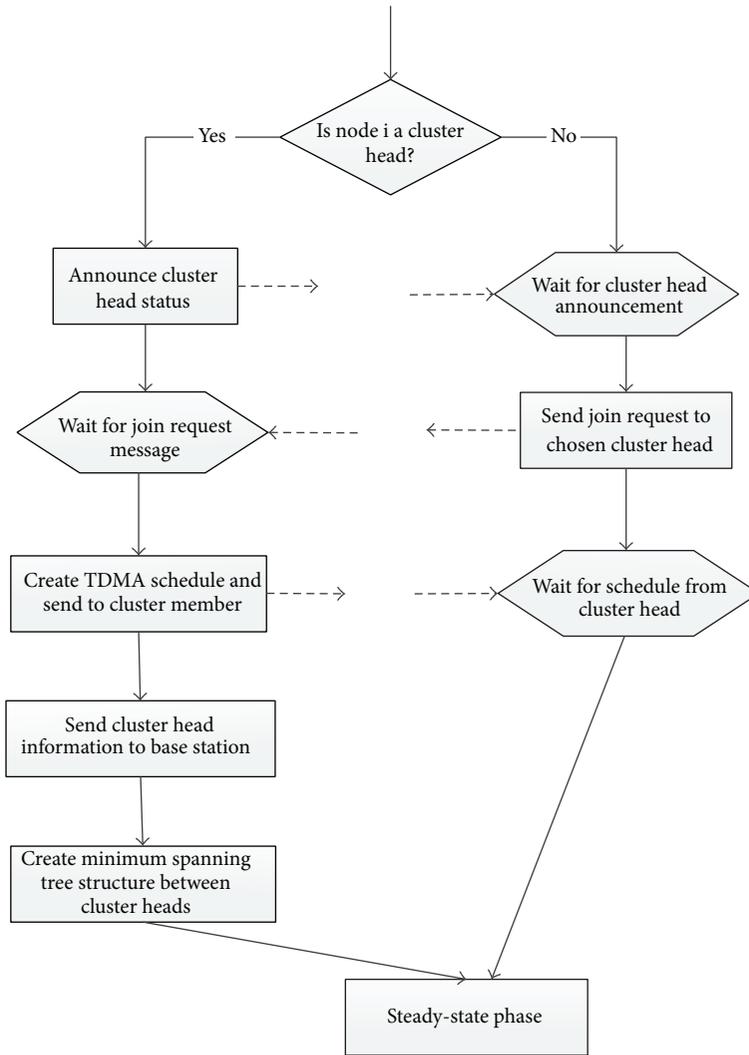


FIGURE 2: Flowchart of steady-state phase in the proposed algorithm, CTDA.

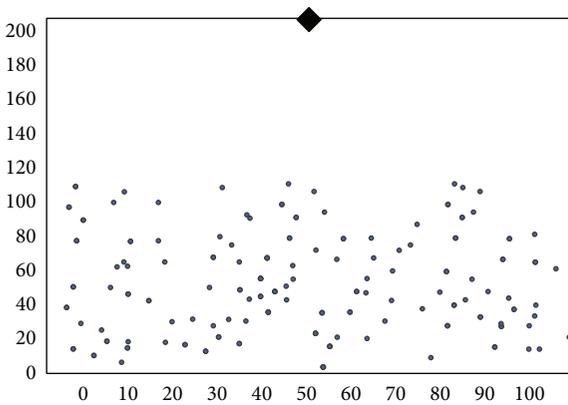


FIGURE 3: Simulated sensor network with 100 randomly distributed nodes (dots) and a base station (diamond on top of the figure).

receives data for the remainder of the experiment; that is, it is considered “dead.”

TABLE 1: Simulation parameters setting.

Parameter	Value
Initial energy	0.5 J
$E_{elec} = E_{TX} = E_{RX}$	50 nJ/bit
ϵ_{fs}	10 PJ/bit/m ²
ϵ_{mp}	0.0013 PJ/bit/m ⁴
EDA (data aggregation)	5 nJ/bit/signal
Control packet size	800
Data packet size	4000

The simulation results in Figure 4 illustrate that the insolubility of nodes (i.e., dead nodes) in CTDA algorithm is more than MCTDA and LEACH protocol.

In addition to reduced node solubility, CTDA algorithm is more energy efficient throughout the simulation. Simulation results are illustrated in Figure 5.

Figure 6 shows the total energy dissipated used by LEACH and other schemas during the simulation network.

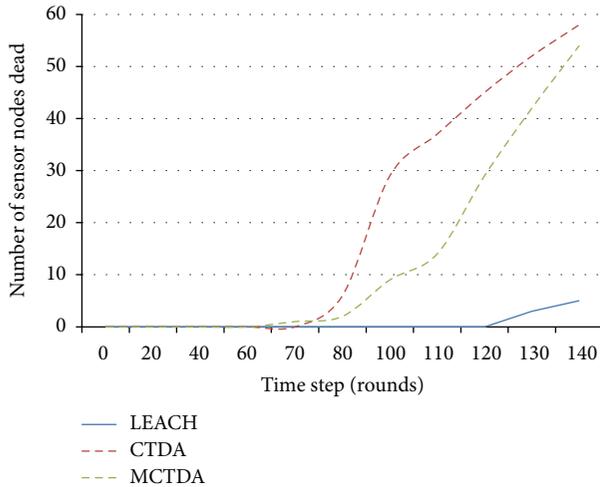


FIGURE 4: Dead nodes during the simulation.

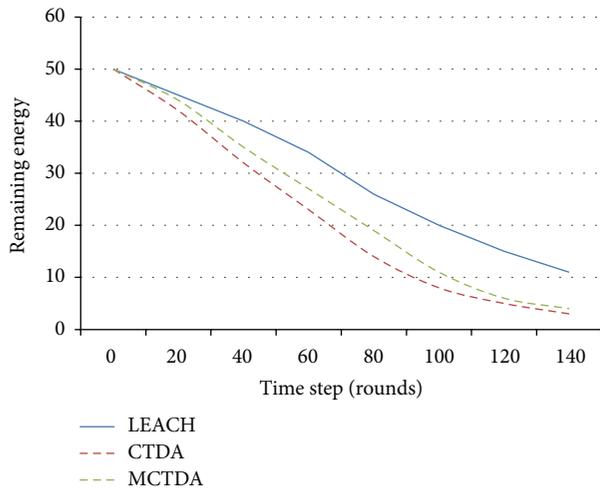


FIGURE 5: Remaining energy of nodes.

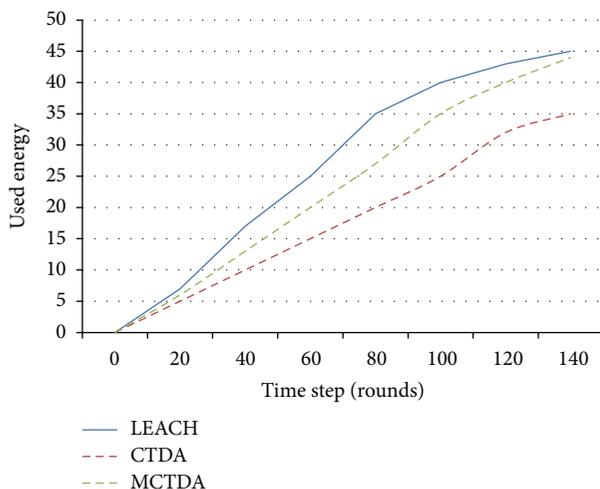


FIGURE 6: Total energy dissipated.

Based on result in Figure 6, minimum cluster head nodes send data to base station, which reduces total energy consumption.

6. Conclusion

Limited energy and redundant data in WSNs require data aggregation communication protocols to minimize the number of sensors that transmit data to the base station. Here, we presented two main protocols in this context including cluster based data aggregation and tree based data aggregation. Cluster based method reduces energy dissipation and collision in a local cluster but energy dissipation for each cluster head to communicate with the base station is high. The tree based method has low distance between each node and its parent, thereby reducing energy dissipation. However, the depth of the tree is high. Here, we proposed an energy efficient hybrid schema of a cluster and tree based data aggregation, CTDA. This proposed hybrid protocol is based on both the LEACH protocol and minimum spanning tree method between the cluster heads. This hybrid schema used advantages of the clustering and tree structures while minimizing disadvantages of each. CTDA consists of two phases, set-up phase and steady-state phase. Comparison of the proposed protocol with LEACH protocol showed that it performs better than LEACH.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research has been supported by a grant from the University of Tehran (no. 29999/01/01).

References

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.
- [2] S. Patil and P. R. Patil, "Data aggregation in wireless sensor network," in *Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research*, 2011.
- [3] 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.
- [4] M. Ding, X. Cheng, and G. Xue, "Aggregation tree construction in sensor networks," in *Proceedings of the IEEE 58th Vehicular Technology Conference (VTC '03)*, vol. 4, pp. 2168–2172, October 2003.
- [5] W. Wang and B. Wang, "Cluster-based and tree-based power efficient data collection and aggregation protocol for wireless sensor network," *Journal of Information Technology*, vol. 10, pp. 557–564, 2011.

- [6] G. Chabra and D. Sharma, "Cluster-tree based data gathering in wireless sensor networks," *International Journal of Soft Computing and Engineering*, vol. 1, pp. 27–31, 2011.
- [7] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless micro-sensor networks," in *Proceedings of the IEEE International Conference on System Sciences*, pp. 1–10, 2000.

