New Data Clustering Algorithm (NDCA)

Abdullah Abdulkarem
Mohammed Al-Matari
Information Technology Department,
Faculty of Computers and
Information,
Cairo University, Cairo, Egypt

Prof. Imane Aly Saroit Ismail
Information Technology Department,
Faculty of Computers and
Information,
Cairo University, Cairo, Egypt

Dr. Amira Mohammed Kotb
Information Technology Department,
Faculty of Computers and
Information,
Cairo University, Cairo, Egypt

Abstract—**Wireless sensor networks (WSNs) have sensing, data processing and communicating capabilities. The major task of the sensor node is to gather the data from the sensed field and send it to the end user via the base station (BS). To satisfy the scalability and prolong the network lifetime the sensor nodes are grouped into clusters. This paper proposes a new clustering algorithm named New Data Clustering Algorithm (NDCA). It takes optimal number of the clusters and the data packets sent from the surrounding environment to be the cluster head (CH) selection criteria.**

Keywords—Wireless sensor network; clustering; energy efficiency; cluster head selection

I. INTRODUCTION

Wireless Sensor Networks (WSNs) is composed of hundreds or thousands of nodes that cooperatively sense the environment to provide interaction between computers or persons and the surrounding environment [1]. A wireless sensor network (WSN) consists of a large number of small sensors with low battery power, low cost, limited storage, a radio transceiver, a tiny microprocessor and a set of transducers [2]. The main task of a sensor node is to gather data from the sensed field and sends it to a common destination called base station (BS). Energy efficiency is the major design goal in these networks. The problem is how to develop a clustering technique that takes into consideration the energy consumption of the sensor nodes. This paper produces a New Data Clustering Algorithm (NDCA) that aims to improve the clustering algorithms in the wireless sensor networks by reducing the energy consumption and maximizing the lifetime of WSNs. The contributions of the NDCA algorithm are the fair divisions of the network area and the change in the cluster head selection criteria.

This paper is organized as follows: Section 2 illustrates the literature review of the clustering algorithms that are used in the wireless sensor networks. Section 3 describes the proposed algorithm in details. Section 4 presents the description of the simulation environment and the performance evaluation of the proposed algorithm. Finally, section 5 presents the conclusion and suggests future work.

II. LITERATURE REVIEW

There are many clustering approaches have been proposed by the research community to address the challenging of WSNs lifetime issue.

The Low Energy Adaptive Clustering Hierarchy (LEACH) protocol is the first attempts in the field of nodes clustering in WSN [3]. The essential idea behind LEACH is a CH rotation role among all the nodes to achieve load balancing. Each sensor elects itself to be a local cluster-head at any given time with a specific probability. Each node generates a random number between 0 and 1, then compares this number with the threshold T(n), if the random number is less than the pre-determined threshold T(n), the node is selected as CH.

The Hybrid Energy-Efficient Distributed clustering (HEED) protocol is a multi-hop clustering protocol for WSNs [4]. It periodically selects CHs based on a hybrid of the node remaining energy and a secondary parameter. The secondary parameters can be the node adjacency to its neighbors or the node degree.

Rotated Hybrid Energy-Efficient and Distributed clustering (R-HEED) protocol developed the HEED protocol by preventing the clustering approach at every round according to certain rules and schedule [5]. At the beginning of every round, CHs wait a period for a re-clustering message from the BS. If they do not receive a re-clustering message, they continue rotating the cluster head role within the same cluster. However, this protocol does not take into account the CH selection criteria which stay randomly rotating.

The Hybrid Energy Efficient Reactive (HEER) protocol is designed to deal with the characteristics of active homogeneous WSNs [6]. In HEER, the Cluster Head (CH) selection is based on the ratio of the residual energy of the node and the average residual energy of the network. Moreover, to conserve more energy, it introduced Hard Threshold (HT) and Soft Threshold (ST) as constraints when the data packets are transmitted over the network.

Another well-known but more efficient hierarchical-based algorithm is LEACH Inner Cluster Election algorithm (LEACH-ICE) [7]. The algorithm selects the cluster head based on the residual energy of the node using the following equation:

$$T_{(n)} = \frac{p}{1 - p[r \bmod (1/p)]} * \frac{E_{n\_resident}}{E_{n\_initial}}, \qquad n \in G \qquad (1)$$

Where $T(n)$ is the threshold of nodes to be selected as CH, P is the probability of the node to be CH, r is the current round, $En\_resident$ is the resident energy of the node, $En\_initial$ is the initial energy of the node and G is the set of

nodes that have not been cluster-heads in the last $\frac{1}{p}$ rounds. Also, the LEACH-ICE algorithm specifies that some cluster head should choose the nearest node inside the cluster to be the new cluster head, when the resident energy of the node is lower than *(∂ Esend)*, where *∂* is a constant value equal to 31.5 and *Esend* is the node's minimum energy level equal to 0.02 Joule.

Distributed Energy-Efficient Clustering (DEEC) protocol, In this protocol, the CH selection criteria is based on the ratio of the residual energy of each node and the average energy of the network [8].

An Efficient Ad-Hoc Routing using a Hybrid Clustering Method in a Wireless Sensor Network algorithm uses a single set-up process to achieve high energy efficiency in wireless sensor networks [9]. It relies on the rotation sequence for selecting a CH instead of the random rotation. The CH node is replaced in each round according to a schedule based on an internal procedure within a sensor node without sending or receiving any additional information.

Energy-Balanced Unequal Clustering (EBUC) protocol is a centralized protocol that organizes the network in unequal clusters and the CHs relay the data of other CHs via multi-hop routing [10]. The operation of clustering is done by the BS. The BS computes the average energy level of each node and uses this information for CH selection.

III. THE PROPOSED ALGORITHM (NDCA)

This paper proposed a new data clustering algorithm named New Data Clustering Algorithm (NDCA).

The nodes are generated randomly and distributed in the network region. The BS gets the nodes locations. Then the network area is divided by the BS to zones that have the same size called clusters. After the nodes have been sensing their environment, some nodes have more events in their field than others. So the nodes have an unequal number of the data packets at a period of time, based on the place of the event. So the minimum number of the data packets is the important factor in this phase. So the node that has the minimum number of packets to send is selected as a cluster head. Also, the residual energy of a selected CH must be higher than the pre-determined threshold. The cluster head is replaced only when its level of the energy falls under a specific threshold. The threshold is computed using formula:

$$Threshold = C * E_{send} \qquad (2)$$

Where *C* is a constant number equal to 31.5, *Esend* is the minimum energy level that enables the node to send packets, which is equal to 0.02 Joule.

Then cluster head broadcasts its unique number to all the nodes that allocated in the same cluster. Fig 1 shows the clustering setup stage of the NDCA algorithm. The NDCA algorithm depends on the LEACH-ICE algorithm with changing in the CH selection criteria. The figure shows the additional steps are used by NDCA over LEACH-ICE.
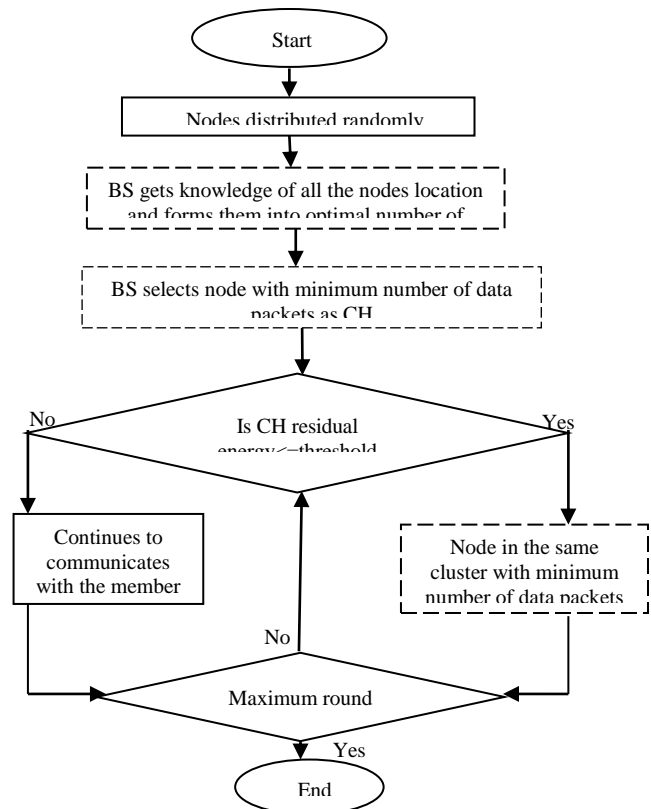

Fig. 1. Proposed NDCA Algorithm Flow Chart

– – – – Additional Steps Compared to LEACH-ICE

IV. SIMULATION AND RESULTS

A simulation program was built by MATLAB for both the NDCA algorithm and the LEACH-ICE algorithm to evaluate their performance and compares between them. The network consists of 100 and 250 nodes. Randomly, nodes deploy in the network into an area of 100m x 100m and 250m x 250m respectively. The base station is located away from the network area and have unlimited resources. The position of the base station (BS) is at location (50,175) and (125,300) respectively. Also it knows the information of the node's ID and location. All nodes have the same capabilities of initial energy, processing resources, and communication. The energy consumption for reception and transmission to all nodes are the same. The nodes states can be active or sleep. The energy consumption for the data transmission depends on the distance between (the sender and the receiver). A node dies when the residual energy is below or equal to 0.02 Joule [7].

Table l lists the used simulation parameters [8, 11].

TABLE I.   SIMULATION PARAMETERS

The Parameter	The Value
Number of nodes	100, 250
Size of network	(100 x 100 m^2), (250 x 250 m^2)
Base Station position	(50 x 175 m^2), (125 x 300 m^2)
Packet size	3500 bits
Control packet	150 bits
Initial energy	2.5 Joule
Threshold of energy depleted	0.02J
Energy dissipated per bit (Eelec)	50nJ/bit
Data aggregation energy consumption (EDA)	5nJ/bit
Distance of threshold (do)	87 m
Amplifier of transmit if $D_{BS} >= do$ (Emp)	0.0013pJ/bit/ m^4
Amplifier of transmit if $D_{BS} <= do$ (Efs)	10nJ/bit/ m^2
Method of deployment	Hybrid

A. Performance Metrics

Five evaluation metrics are used to evaluate and test the performance of the proposed NDCA algorithm and compare it to the LEACH-ICE algorithm.

1) Average Energy Consumption

It is the average energy consumption for all the alive nodes of the whole network.

$$\text{Average Energy Consumption} = \sum_1^n \frac{\text{energy consumption of nodes}}{\text{Number of alive node (n)}}$$

(3)

It is better to keep it as a least as possible.

2) Number of Alive Nodes per Round

It is the lifetime interval of the nodes between the network operation start until the death of the last node. It is computed as follows:

$$\text{Average number of alive nodes} = \sum_1^n \frac{\text{Number of alive nodes}}{\text{Number of round (n)}} \quad (4)$$

It is better to maximize the metric.

3) Number of Various Nodes Dying per Round

a) FND (First Node Dies): It is the first node that has depleted its whole energy (specified threshold) in a round during the network life time.

b) HND (Half of the Nodes Die): It is the rounds (time) elapsed until half of the nodes have consumed their whole energy (specified threshold).

c) LND (Last Node Dies): It is the last node that has exhausted its whole energy (specified threshold) in a round during the network life time.

It is better to maximize the metric.

4) Number of Packets Delivered to Cluster Head

It is the number of packets that the CHs received over the network lifetime.

Average Packets Delivered to CH =

$$\sum_1^n \frac{\text{Number of packets reached to CH}}{\text{Number of round (n)}} \quad (5)$$

It is better to maximize the metric.

5) Number of Packets Delivered to Base Station

It is the total number of packets that the base station received during the network lifetime.

Average Packets Delivered to BS =

$$\sum_1^n \frac{\text{Number of packets reached to BS}}{\text{Number of round (n)}} \quad (6)$$

It is better to maximize the metric.

B. Evaluated Scenarios

A simulation program was built for two scenarios. In the first scenario the network size is 100 nodes in (100m x 100m) area and the base station location is at (50m x 175m). While in the second scenario the network size is 250 nodes in (250m x 250m) area and the base station location is at (125m x 300m).

1) First Scenario

In this scenario; the network size is 100 nodes distributed over (100m x 100m) area and the base station location is at (50m x 175m).

a) Average Energy Consumption

Fig 2 shows the energy consumption for both NDCA and LEACH-ICE algorithms. Ten rounds have been extracted randomly from along the rounds range, in order to compare the proposed NDCA with LEACH-ICE algorithms. It is obvious from the figure that the NDCA algorithm saves energy more than that of the LEACH-ICE algorithm by 27%. This energy conservation is due to the criteria of CH selection which chooses the node that has the minimum packets to send to be selected as CH. Unlike LEACH-ICE which it is based on the random distribution of CHs and selection criteria.
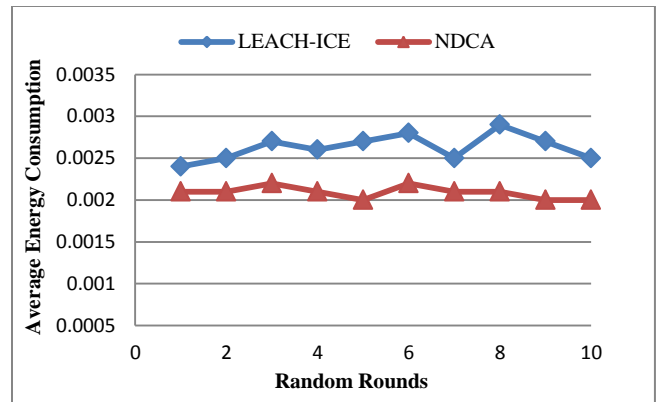

Fig. 2.   Average Energy Consumption

b) Number of Alive Nodes per Round

Fig 3 shows the number of alive nodes per rounds for the NDCA and the LEACH-ICE algorithms. It is obvious from the figure that the proposed NDCA algorithm achieves better performance than that of the LEACH-ICE algorithm. The proposed algorithm extended the network life-time more than LEACH-ICE by 272 rounds, which is about 16%. This improvement is due to the energy conservation during the clusters setup phase achieved, as while LEACH-ICE treats all the nodes without discrimination according to the data load of a node, but NDCA takes data load of nodes into account; therefore, has a longer period of the network lifetime than LEACH-ICE.
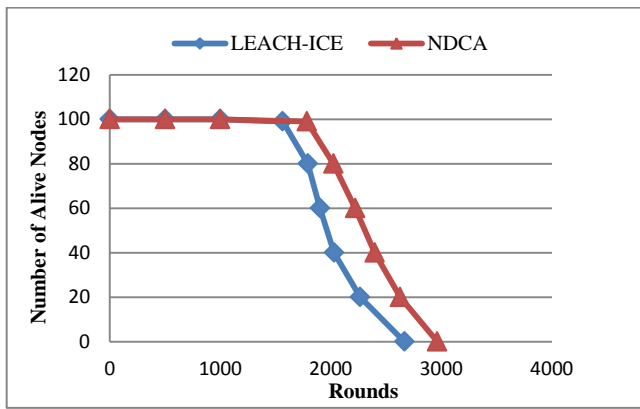
Fig. 3. Number of Alive Nodes per Round

c) Number of Various Nodes Dying per Round

Fig 4 demonstrates the performance measurement in term of the number of various of nodes that die per round for NDCA and LEACH-ICE algorithms. The graph shows that the lifetime of FND metric in the NDCA algorithm compared to the LEACH-ICE algorithm increased by 14%, HND metric by 16%, and LND metric by 10% respectively. Since NDCA optimize energy consumption of nodes communication, as it keeps the cluster head with a light load, high residual energy and with fair distribution which leads node to be close to the cluster head, this has a vital impact on the lifetime of the nodes in the network.
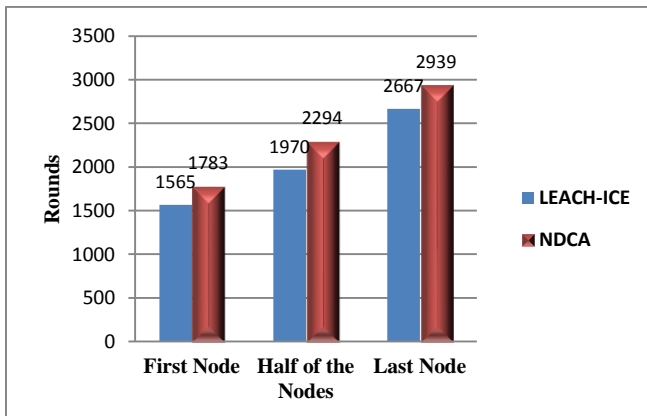

Fig. 4. Number of Various Nodes Dying per Round

d) Number of Packets Delivered to Cluster Head

Fig 5 shows the number of packets delivered to the CH for the NDCA and LEACH-ICE algorithms. The figure obviously shows that the data packets transmitted over throughout the network to CHs by NDAC are more than that transmitted by LEACH-ICE. The proposed algorithm achieved better performance than that of the LEACH-ICE algorithm by 16% in this term. This achievement is due to the increase in the network lifetime produced by the process of the cluster heads distribution and selection. The number of data packets sent to the CHs is slightly more than that of LEACH-ICE because the node with a minimum number of the packet is selected as CH, so others nodes with a large number of packets send to the CH.
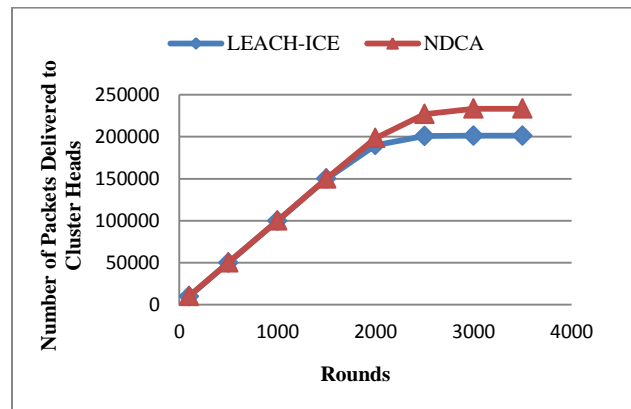

Fig. 5. Number of Packets Delivered to Cluster Heads

e) Number of Packets Delivered to Base Station

Fig 6 shows the number of packets delivered to the BS for the NDCA and LEACH-ICE algorithms. It is clear from the figure that the data packets transmitted through the network from the CHs to the BS by the NDAC algorithm are more than that transmitted by the LEACH-ICE algorithm. The proposed algorithm outperforms the LEACH-ICE algorithm by 18% in this term. Since in NDCA, CHs received much more data packets than that in LEACH-ICE as mentioned in the above subsection, cluster heads periodically aggregate this data and then send them to the BS.
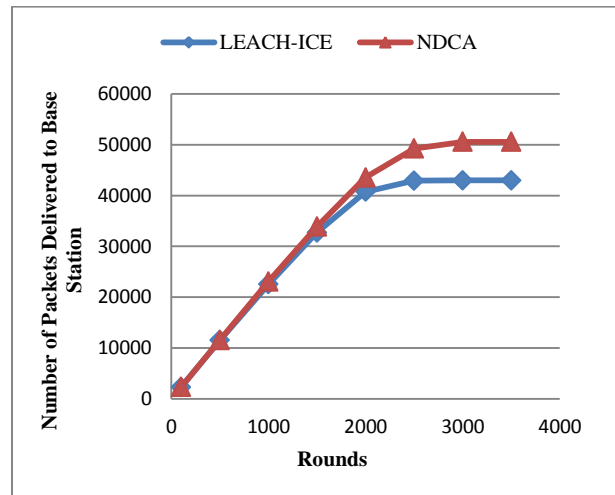

Fig. 6. Number of Packets Delivered to Base Station

2) Second Scenario

In this scenario the network size is 250 nodes distributed over (250m x 250m) area and the base station location is at (125m x 300m).

a) Avergae Energy Consumption

Fig 7 displays the energy consumption for both NDCA and LEACH-ICE algorithms. Fig 7 shows that the NDCA algorithm consumes less energy than that of the LEACH-ICE algorithm; therefore, the NDCA algorithm saves energy by 38%. This improvement is due to the fact that the LEACH-ICE algorithm consumes more energy than NDCA because of the

unfair distribution of CHs that causes a large distance between the node and the CH. Also, CH selection criteria plays a significant role in prolonging the lifetime of the network.
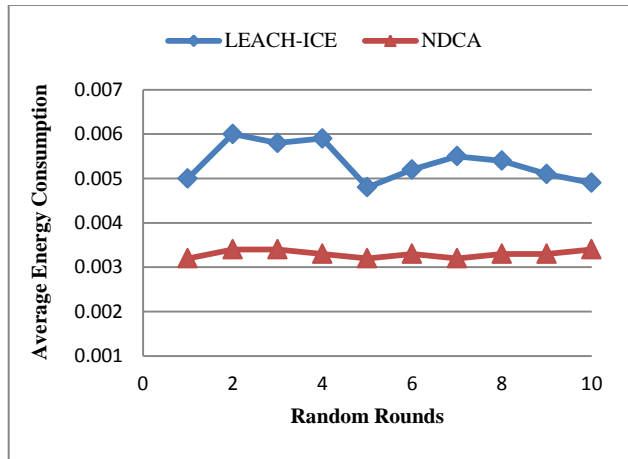

Fig. 7.   Average Energy Consumption

b) Number of Alive Nodes per Round

Fig 8 shows the number of the alive nodes per rounds for the NDCA and LEACH-ICE algorithms. It is clear from the figure that the proposed algorithm saves great amount of the energy in this scenario. Therefore, the NDCA algorithm achieves better performance than that of the LEACH-ICE algorithm by prolongs the network life-time by 411 round, which is about 26%. While in the first scenario the proposed algorithm extends the network life-time by 272 round. This improvement is due to the selection of cluster head based on the minimum data packets to send. Also the residual energy of a node is taken into consideration.
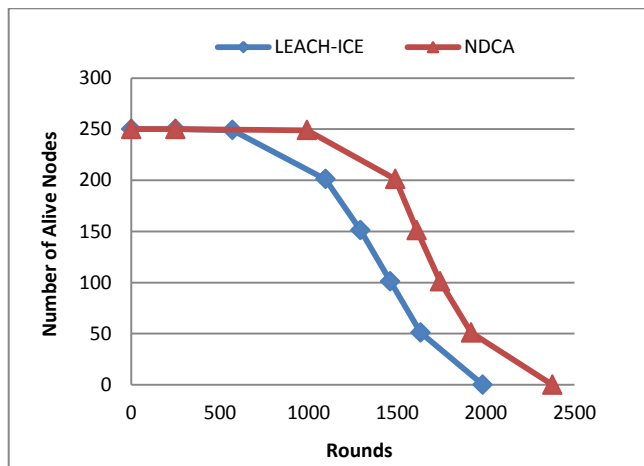

Fig. 8.   Number of Alive Nodes per Round

c) Number of Various Nodes Dying per Round

The numbers of various nodes dying per round for the NDCA and LEACH-ICE algorithms are illustrated in fig 9. It is obvious from the figure that the NDCA algorithm achieves better performance than that of the LEACH-ICE algorithm by 73% for the FND, by 23% for the HND and by 20% for the LND. The increase achievement percentage compared to the first scenario is by 59%, 7% and 10% respectively. This

enhancement is also due to the selection of cluster head based on the minimum data packets to send and the residual energy parameter of the nodes.
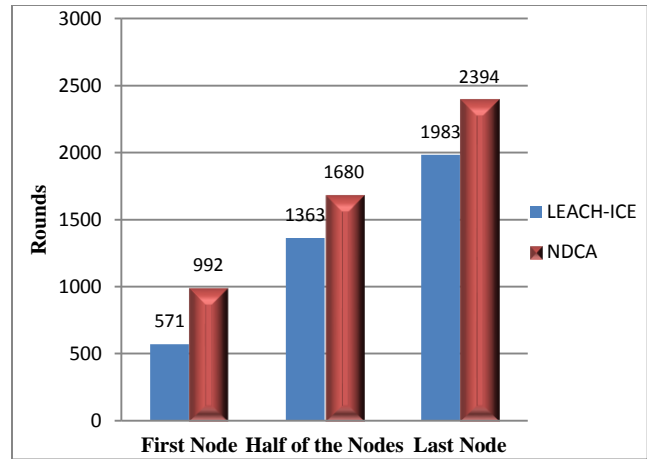

Fig. 9.   Number of Various Nodes Dying per Round

d) Number of Packets Delivered to Cluster Head

Fig 10 demonstrates the number of packets delivered to CH for the NDCA and the LEACH-ICE algorithms. It is observed from the figure that the total packets sent by the proposed NDCA algorithm to the cluster heads in each cluster are much more data packets than that of the LEACH-ICE algorithm. The proposed NDCA algorithm overcomes the LEACH-ICE algorithm in this metric by 26%. The difference in the amount of data packets send to CHs between the first scenario and the second scenario is by 10%. This is because, here the nodes that forward the data packets have huge number of packets, while the CH who received the data packets are selected based on the minimum number of data packets to send, this saves energy for prolonged lifetime. On other hand; in NDCA, the nodes are alive for longer time compared to LEACH-ICE.
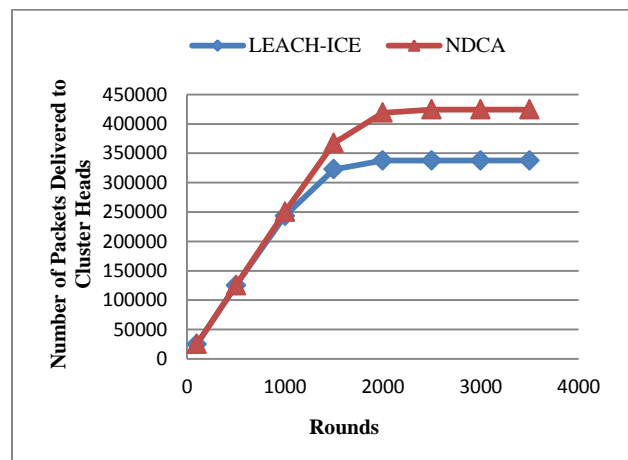

Fig. 10. Number of Packets Delivered to Cluster Heads

e) Number of Packets Delivered to Base Station

Fig 11 shows the number of packets delivered to BS for the NDCA and LEACH-ICE algorithms. Fig 11 displays that the packets received at the base station by the NDCA algorithm are much higher than that received by the LEACH-ICE algorithm.
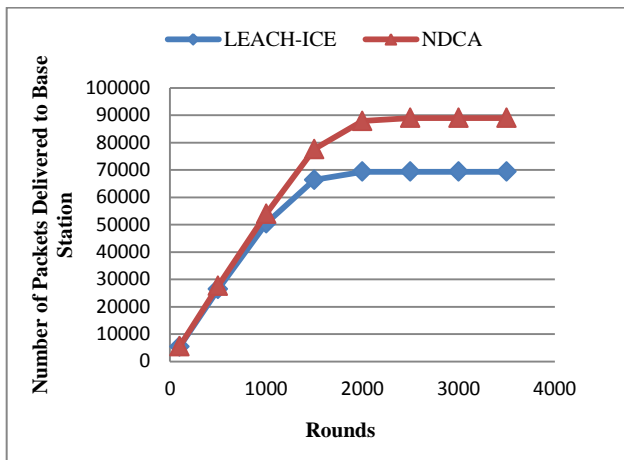
Fig. 11. Number of Packets Delivered to Base Station

In this term, the result of the proposed NDCA algorithm improves the LEACH-ICE algorithm by 28%. Better network cluster head distribution and efficient cluster head selection criteria are two main reasons of increased the data packets in the NDCA algorithm.

3) General Comparison

Table 2 contains summarization of the implementation results and the comparison of the algorithms; LEACH-ICE and the proposed NDCA. From the table, It can be observed that the average values of all the performance metrics when applying the proposed NDCA algorithm overcome those when applying the LEACH-ICE algorithm in the all cases.

TABLE II. SUMMARIZES THE COMPARISON BETWEEN NDCA AND LEACH-ICE USING TWO SCENARIOS

Performance Matric	Average Percentage of Improvement Using NDCA	
	First Scenario	*Second Scenario*
Average Energy Consumption	27%	38%
Average Alive Nodes	16%	26%
First Node Die	14%	73%
Half of the Nodes Die	16%	23%
Last Node Die	10%	20%
Number of Packets Delivered to Cluster Head (CH)	16%	26%
Number of Packets Delivered to Base Station (BS)	18%	28%

V. CONCLUSION AND FUTURE WORK

In this paper, a clustering algorithm, named NDCA, for WSN is proposed. This algorithm chooses the node with minimum data packets to send to be as CH and the residual energy of the new CH must be higher than that of the predefined threshold level. Also the proposed algorithm divided the network into zones (clusters). A simulation program was built for both the NDCA algorithm and the LEACH-ICE algorithm to evaluate their performance and compare between them. Two different scenarios are used to evaluate the proposed NDCA algorithm. It is concluded from the results, that the proposed NDCA algorithm outperforms the LEACH-ICE algorithm on all the performance metrics.

For future work, there are many important subjects such as: introducing the clustering technique for Wireless Sensor Network over Cloud Computing. It is important to add an intelligent method for secure communications in this filed.

REFERENCES

[1] R. Verdone, D. Dardari and G. Mazzini, "Wireless Sensor and Actuator Networks," Elsevier, London, UK, 2008.

[2] R. Badi, P. Suri and S. Gupa, "Review Paper on Various Clustering Protocols Used In Wireless Sensor Network (WSN)," Electronics, Signals, Communication and Optimization (EESCO) International Conference, Visakhapatnam, pp. 1- 4, Jan 2015.

[3] W. Heinzelman, A. Chandrakasan and H. Balakrishnan "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," Proceedings of the 33rd Hawaii International Conference on System Sciences, Maui , HI , USA, IEEE, 4-7 Jan 2000.

[4] O. Younis and S. Fahmy ," A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad hoc Sensor Networks," Mobile Computing, IEEE Transactions, vol. 3, no. 4, pp. 366-79, 2004.

[5] W. Mardini, M. B. Yassein, Y. Khamaysh and B. Ghaleb, "Rotated hybrid, energy-efficient and distributed (R-HEED) clustering protocol in WSN," WSEAS Trans. Comm, vol. 13, pp. 275–290, 2014.

[6] S. Mohammed, K. Latif and U. Qasim, "HEER: Hybrid Energy Efficient Reactive protocol for Wireless Sensor Networks," Electronics Communications and Photonics Conference (SIECPC), 2013 Saudi International, Riyadh, pp. 1 – 4, April 2013.

[7] Y. Miao, "Cluster-head Election Algorithm for Wireless Sensor Networks based on LEACH Protocol," Trans Tech Publications, Switzerland, vols. 738-739, pp. 19-22, 2015.

[8] L. Qing, Q. Zhu and M. Wang, "Design of a Distributed Energy-Efficient Clustering Algorithm for Heterogeneous Wireless Sensor Networks," ELSEVIER, Computer Communications 29, pp. 2230- 2237, 2006.

[9] D. Nam and H. Min, " An Efficient Ad-hoc Routing using a Hybrid Clustering Method in a Wireless Sensor Network," Wireless and Mobile Computing, Networking and Communications, Third IEEE International Conference, New York, USA, p. 60, Oct 2007.

[10] J. Chang, S. Wei, X. Min and T. Xian, "Energy-Balancing Unequal Clustring Protocol for Wireless Sensor Network," The Journal of Chain Universities of Posts and Telecommunications, vol.4, p. 94–99, August 2010.

[11] R. N. Enam, S. Misbahuddin and M. Imam, "Energy efficient round rotation method for a random cluster based WSN," Collaboration Technologies and Systems (CTS), International Conference, Denver, vol. 8, p. 157 – 163, May 2012.