Research Journal of Applied Sciences 11 (10): 1122-1129, 2016
ISSN: 1815-932X

Nodes Clustering Approach to Improve the Data Transmission in WSNs

Saad Talib Hasson and Hawraa Abd
College of Information Technology, University of Babylon, Hillah, Iraq

Abstract: "Wireless Sensor Networks" (WSNs) representa collection of many deployed sensors utilized to collect and transfer information. WSNs have many vital applications. WSNs have received increasing consideration in this era due to their useful and vital applications. The main objective of this study is to simulate, evaluate and observe the behavior of a developed clusteringapproach and compare its performance metrics with the DB-scan algorithmin a suggestedWSNs. In this study all the cluster sensor nodes must sensing certain data and must transmit it. These data will be collected at particular nodes known as Cluster Heads (CH) that is previously assigned for each cluster. The CH aggregates the data and forwards it to the base station or a node sink. Many performance metrics can be used to Measurethe performance of the suggested WSN such as NRL, PDF, End-to-end and throughput. Evaluation of the developed clustering algorithm implementation shows certain improvement percentage compared to DB-scan algorithm. Net Logo (5.2.1 version 2015) programming language is suggested to be the multi agent modeling and simulation approach in this study.

Key words: WSN, data transmission, clustering, performance evaluation, simulation

INTRODUCTION

"Wireless Sensor Networks" (WSNs) usually composed of manydeployedunits called sensors. WSN is containing a large number of sensor nodesrandomly deployed in certain area. Sensor nodes are small in size with limited power battery, limited processing and computational resources. The sensors are mainly produced to monitor certain physicalconditions andto collect data such as temperature, light, motion, pressure, humidity and so on. The collected data must be transmitted to a base station or a sink node. The main challenges in WSNs are the network scalability, fault tolerances, security, reliable data transmission, robustness and need for efficient energy.

WSNs must have an ability to work in environments without any human access direct monitoring. In most of the WSNs applications, their sensor nodes are often randomly deployed. Grouping the deployed sensor nodes into clusters has been widely implemented by many researchers. Clustering process supported the "network scalability", helps in localizing the route inside the cluster to decrease the dimensions of the built routing table kept inside one of the node units. It also helps in reducing the energy consumption by data aggregation. Clustering has a major effect in stabilizing the WSN topology. In clustering environment, Sensors in each cluster (called members) will connect only with their Chs (Soni, 2012). Clustering plays a major role in developing the data gathering process and ensuring possible "data fusion" and aggregation. In WSNs each cluster has a leader identified by Cluster Head (CH) and somejoint sensor nodes as members. Sensor nodes are grouped together to form small clusters and a Cluster Head (CH) for each cluster is elected. In this architecture, sensor node transmits data to their respective CH and CH aggregates data and forward them to a base station. Sensor nodes in clusters transmit messages over a short distance within their respective clusters therefore minimum amount of energy exhausted from sensor nodes in clusters but in case of CHs more energy is drained due to message transmission over long distance i.e. CHs to the base Station (Priya and Chouhan, 2014).

MATERIALS AND METHODS

Clustering: Clustering represents the process ofpartitioningnodes into groups, each group called cluster. Many clustering algorithms were developed and implemented in different fields. Most of these are; Hierarchical clustering, Partitioned clustering, density based clustering, grid-based clustering and model-based clustering (Soni, 2012).

Clustering increases the network scalability and its life. It makes scattering control over the network more various. It saves energy by dispensing load by creation intelligent choice. Nodes having high energy are allocated more loads thus increasing the lifetime of the network. The clustering is complete in such a way that data has to portable minimum. Only cluster heads communicates with

Corresponding Author: SaadTalibHasson, College of Information Technology, University of Babylon, Hillah, Iraq

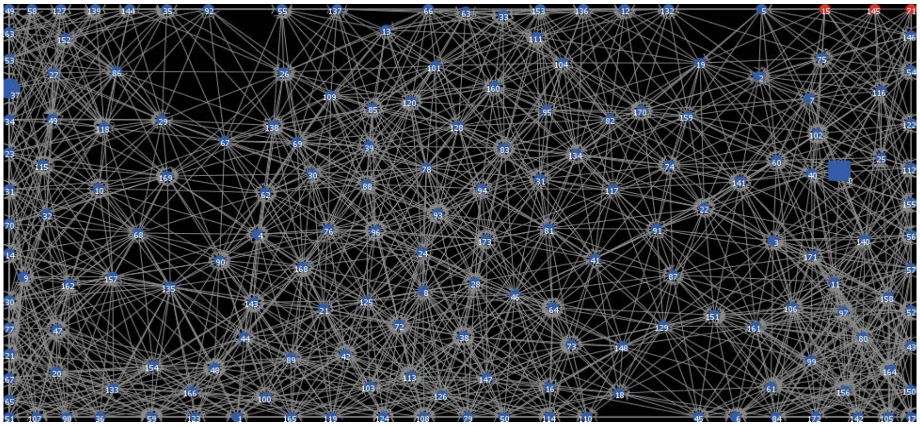

Fig. 1: DB-Scan clustering simulation program

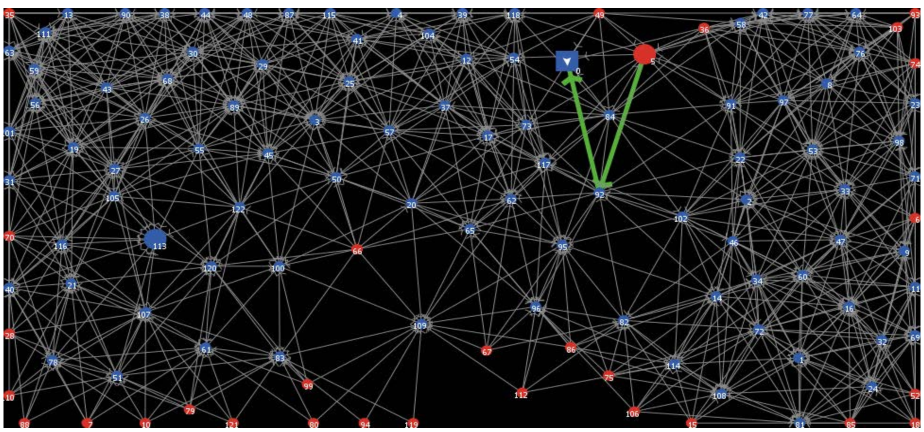

Fig. 2: Data transmission process DB-Scan clustering approach

cluster head thus decreasing the data redundancy which usually occurs when each node implement its own data combination and transmission function individually. Clustering in WSN network makes them appropriate for use in irregular environments (Mamta, 2014).

DB-Scan algorithm: DB-Scan ("Density Based Spatial Clustering Applications with Noise") is a simple density based clustering algorithm suggested by Martin *et al.* (1996). They aimed to discovertherandom shaped clusters in addition toseparate noise from large spatial databases. It considers two essential concerns namely "Eps (radius)" and "MinPts (minimum points-a threshold)". A density represents the number of points inside the known radius. The density is estimated by calculating the number of points within a specified radius Eps, basing ona "center-based approach". Any point can be classified as a border point, a core point or a noise point. The neighbor of a certainrangemust hold at least a smallest number of points. It identified clusters in big data collections by observing its local elements density depending on one input parameter. Minimum number of points (Minpts) used to decide whether about the dense of a neighborhood while the (Minpts) identify the density threshold of the evener regions.it can discover clusters with arbitrary shapes and size (Soni, 2012).

Net Logo (5.2.1) was used in building a suitable environment to simulate and apply DB-Scan algorithm for the created network. Figure 1 showsa snapshot of this simulation program.

To apply the DB-Scan algorithm we distribute the nodes randomly in asuggested environment at the first step. The DB-Scan algorithm was implemented, observedand operated to assign the clusters, CHs, their members and transmitting data. Figure 2 shows a sample of the data transmission process from a source node to the sink node.

Developed clustering approach: To improve the process of data transmission in WSNs, a developed clustering

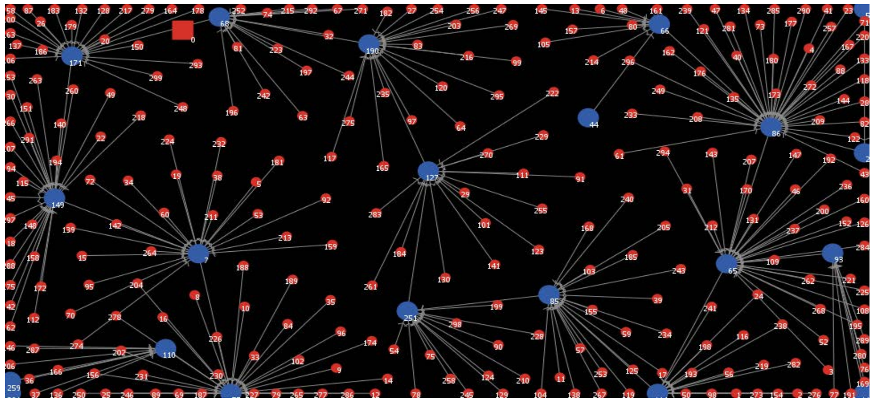

Fig. 3: Selecting the CHs and members in the proposed clustering approach

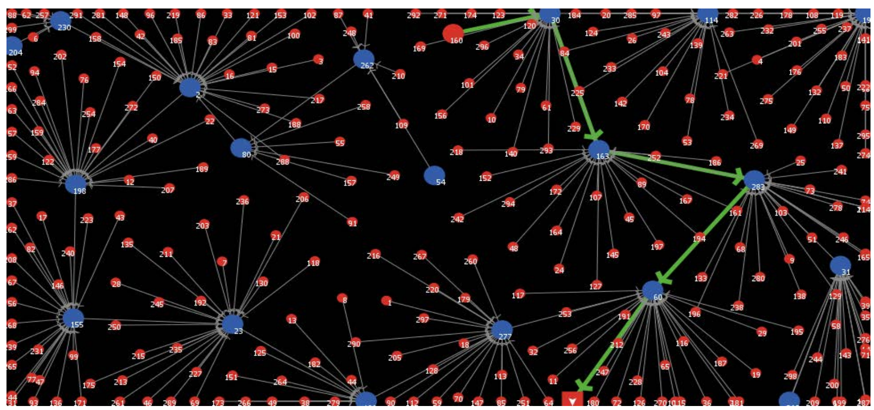

Fig. 4: Transferring data to the BS in the proposed clustering approach

approach was built and simulated using Net Logo (5.2.1). A network main Base Station (BS) location must be indicated and identified initially. Its shape, size and colour are designed to observe and recognize it visually. The program interface offers a facility to select the BS location either manually by the user or randomly by the program. The proposed approach suggestedthe process of selecting the first Cluster Head (CH) as a program input either by the user interfaceor randomly by the program depending on the BS location. The CH node must be identified by (CH) its colour and size must initially be assigned also. Let all the nodes within the CH transmission range as cluster members in this first cluster. These nodes will be identified as (m) their colours and sizes must also be selected. Select the far node in this cluster to be the next CH and change its properties. Starting from this new CH node, a new check will be done to select its member nodes by collecting all the nodes in the transmission range of the selected cluster head and not joined any other cluster. This process will continue with all other network nodes.

In the developed clustering approach different number of sensor nodes was distributed randomly in a suggested environment at the first step. The next step is to select the first cluster-head which can be achieved either by the user or randomly. Then to selectall the nodes in its transmission range to be cluster members to form the first cluster. In order to ensure and success the process of sending a data to the sink node, we must select one of the members which have a maximum distance to the CH to be the next cluster head and so on. Every member node send data only to its clusterhead of its cluster but the clusterhead can send data to any neighbor clusterhead within the range. Figure 3 show the process of selecting the CHs and member nodes during the implementation of a suggested simulation program depending on the proposed clustering approach using Net Logo.

After completing the process of clustering and creating other network parameters, an implementation was applied. Simulating this network in many runs will give a good estimation about its behavior. Figure 4 show the process of transfer data to the sink through proposed clustering approach by using the Net Logo simulator.

Performance evaluation: There are several metrics can be used inanalyzingand evaluating the performance of the WSNs.Some of these metrics are (Saad, 2013).

The throughput: The throughput was defined to be the amount of received data by thedestination nodes duringa period of time (Ducksoo *et al*., 2009).

Average "end-to-end delay": This metric was defined to be the required time for the packet to be delivered by the sink node.It can beestimated by accumulating the time of all the received packets to the number of all packets (Alexandre, 2004).

"Packet Delivery Fraction" (PDF): PDF represents the percentage of the received to the transmitted packets (Mamta, 2014).

$$\text{PDF} = (\text{Received number}/\text{Total number sent})\times 100$$

"Normalize Routing Load" (NRL): NRL represents the ratio of the transmitted packets to the received packets (Alexandre, 2004).

$$\text{NRL} = \text{No. of transmitted packets}/\text{No. of the received packets}$$

"Dropped packets": Any sent packet and fail to be receipt by the destination called dropped. It indicates the number of the failed packetsor the subtraction of the received from the transmitted packets (Aliff *et al*., 2006).

RESULTS AND DISCUSSION

Simulation results: There are many parameters effects the process of transferring messages in WSNs. In this study; the effects of nodes numbers and transmission range (coverage area of the sensor node)on the WSN behaviour were observed and evaluated in two cases.

Case 1: The first scenario a developed clustering approach was implemented. The effect of varying nodes number (ranges from 50-500) nodes was simulated and evaluated. Eachscenario was run and simulated in an operation manner 30 time in order to get close real results from the simulation program. The resulted information was listed and averaged. Table 1 shows the built environment parameters in this simulation scenario set up. This scenario suggested that all parameters are fixed except the numbers of nodes are variable. The simulation program with specific number of nodes was repeated 30 times and then organized the final simulation results in a large data tables. Table 2 shows the final average results.

In the second scenario the same environment was simulated, observed and implemented but with a developed clustering algorithm. Table 3 shows the final average results after these simulation runs for the developed clustering algorithm.

Table 1:WSN environment of case1

Parameter	Value
The simulator	Net Logo 5.2.1 version (2015)
Nodes number	Change able (50,100,200,300 and 500)
Algorithm	Developed clustering
Range	15 m
Pause time	5s

Table 2: Simulation results for the first scenario of case1

Nodes number	Average received messages	Average lost messages	Average NRL	Average PDF	Average End-to-End	Average through put
50	1	0	0.576667	57.66667	0.753333	0.458889
100	1	0	0.7	70	0.9	0.566667
200	1	0	0.616667	48.33333	0.766667	0.4
300	1	0	0.783333	63.33333	0.9	0.555556
500	1	0	0.79	57.66667	0.94	0.476667

Table 3: Simulation results of case1 using developed clustering algorithm

Nodes number	Average received messages	Average lost messages	Average NRL	Average PDF	Average End-to-End	Average through put
50	1	0	0.85	62.5	1.04	0.528333
100	1	0	0.725	68.33333	0.9	0.575
200	1	0	0.708333	59.16667	0.866667	0.494444
300	1	0	0.666667	66.66667	0.833333	0.555556
500	1	0	0.666667	66.66667	0.833333	0.555556

The results in Table 2 and 3 were graphed in Fig. 5-8 to show the relationship between the network metrics and their effects on it's behavior. Figure 5 show the relationship betweenaverage NRL in two algorithms (DB-Scan and developed clustering) and the number of nodes.

Figure 5 shows that the average NRL of the developed clustering is less than the average NRL of the DB-Scan at all the number nodes except when the node number is (300 and 500). Figure 6 shows the relationship between average PDF in two clustering algorithms (DB-Scan and developed clustering) and number of nodes.

Figure 6 shows that the average PDF of the developed clustering is in general less than the average PDF of the DB-Scan at all the number nodes. Figure 7 shows the relationship between average End-to-enddelay in the two clustering algorithms and the number of nodes.

Figure 7 shows that the average End-to-end of the developed clustering less than average End-to-end of the DB-Scan in small number of nodes (<250) while the average End-to-end of the developed clustering is being more than average End-to-end delay of the DB-Scan when the number of nodesgreater or equal to 300 nodes. Figure 8 shows the relationship between average throughput in the two algorithms and the number of nodes.

Figure 8 shows that the average throughputvalues of the developed clustering are less than thatof the DB-Scan in all number of nodes.

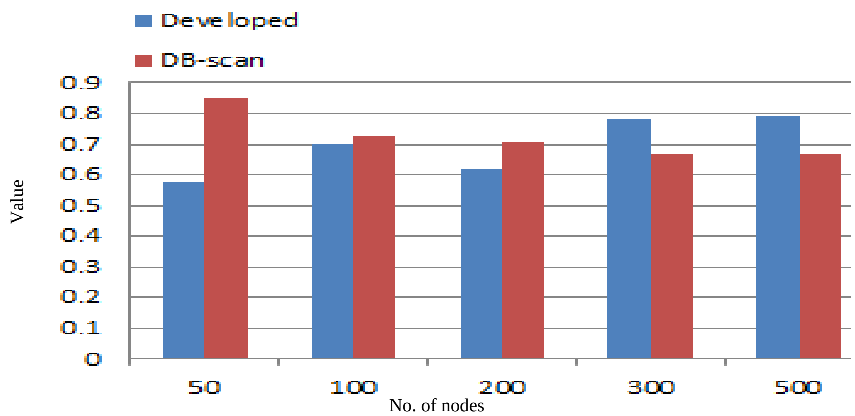

Fig. 5: NRL values with the number of nodes in the two clustering approaches

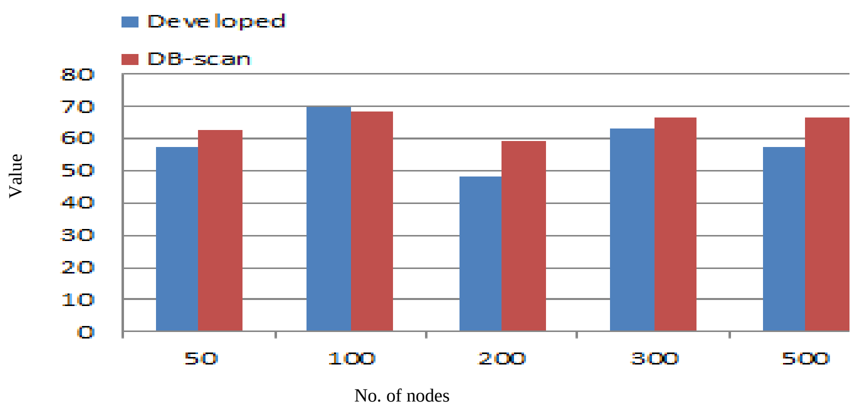

Fig. 6: PDF values with the number of nodes in the two clustering approaches

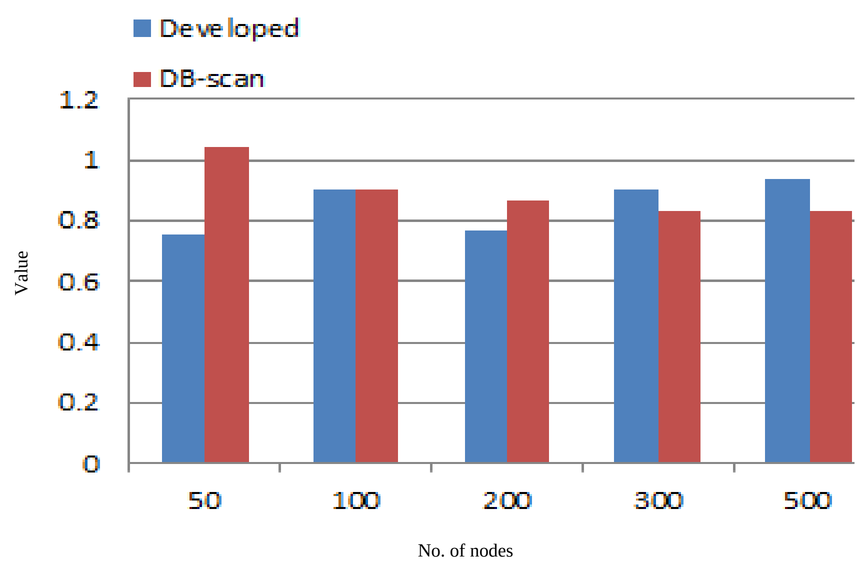

Fig. 7: End-to-End delay values with the number of nodes in the two clustering approaches

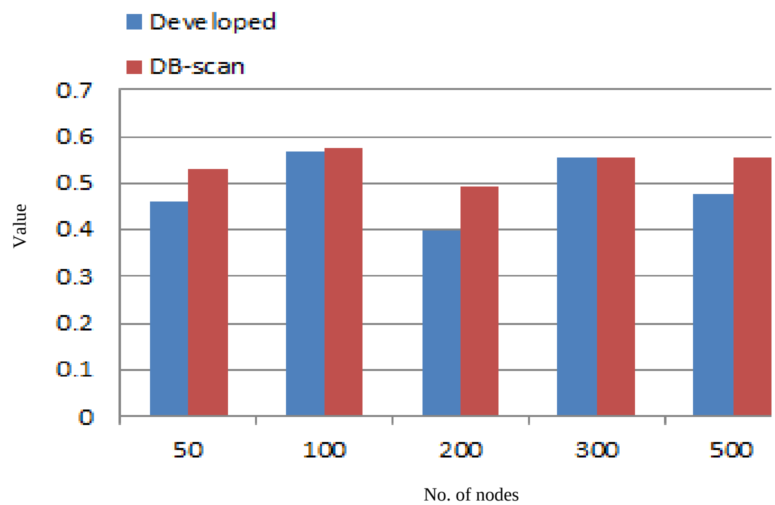

Fig. 8: Throughput values with the number of nodes in the two clustering approaches

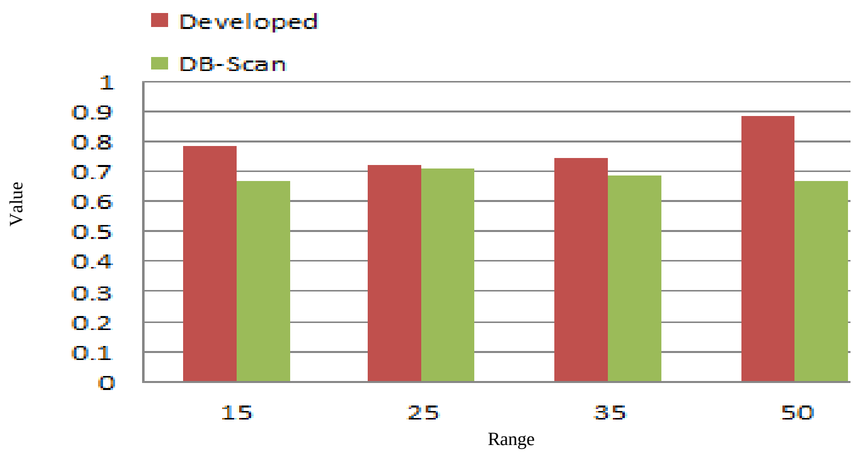

Fig. 9: Effects of different transmission ranges on NRL values

Case 2: The secondscenario represents a simulation approach to compare the behaviour of WSNs in two different clustering applications. The effect of varying the nodes "transmission range"or the "coverage area"(ranges from 15-50) was applied and evaluated in many simulation runs. Each scenario was run and simulated 30 times in order to get close real results from this simulation program. The resulted information was listed and averaged. Table 4 shows the built environment parameters in this simulation scenario set up.

This scenario suggested that all the parameters are fixed except the transmission range of nodes is variable. Thefinal simulation results for the performance evaluation metrics were collected and averaged. Table 5 shows the final average results.

Table 6 shows the final average simulation results for the DB-Scan clustering approach. The results in the Table 5 and 6 were graphed in Figs. 9-12. These figures show the relationship between the nodes coverage area

Table 4:WSN environment setup for case2

Parameter	Value
The simulator	Net Logo 5.2.1 version (2015)
Nodes number	300
Algorithm	Clustering
Range	15, 25, 35, 50 m
Pause time	5s

Table 5: Case 2 Simulation results for the developed clustering

Range	Average received messages	Average lost messages	Average NRL	Average PDF	Average End-to-End	Average through put
15	1	0	0.783333	63.33333	0.9	0.555556
25	1	0	0.721429	37.14286	0.832857	0.310476
35	1	0	0.741667	62.5	0.833333	0.572222
50	1	0	0.885	62.66667	1	0.571667

Table 6: Case 2 Simulation results forthe DB-Scan algorithm

Range	Average received	Average lost	Average NRL	Average PDF	Average End-to-End	Average throughput
15	1	0	0.666667	66.66667	0.833333	0.555556
25	1	0	0.708333	59.16667	0.866667	0.494444
35	1	0	0.683333	68.33333	0.866667	0.561111
50	1	0	0.666667	66.66667	0.833333	0.555556

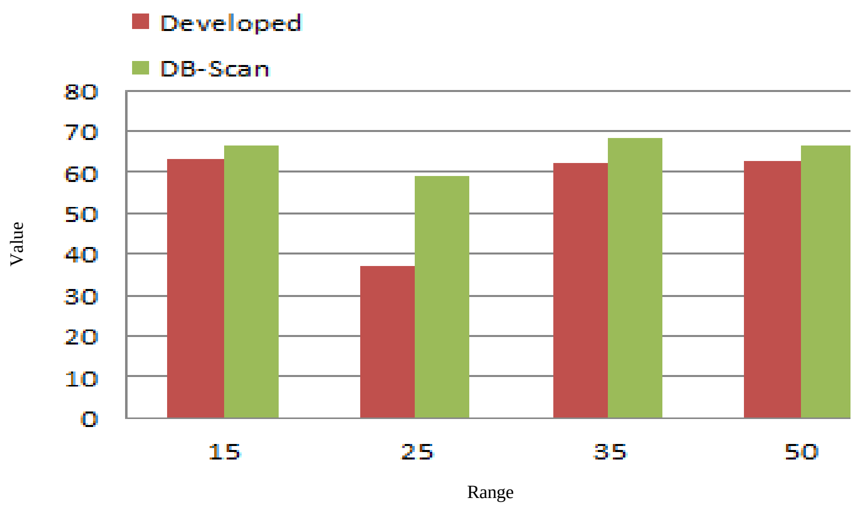

Fig. 10: Effects of different transmission ranges on PDF values

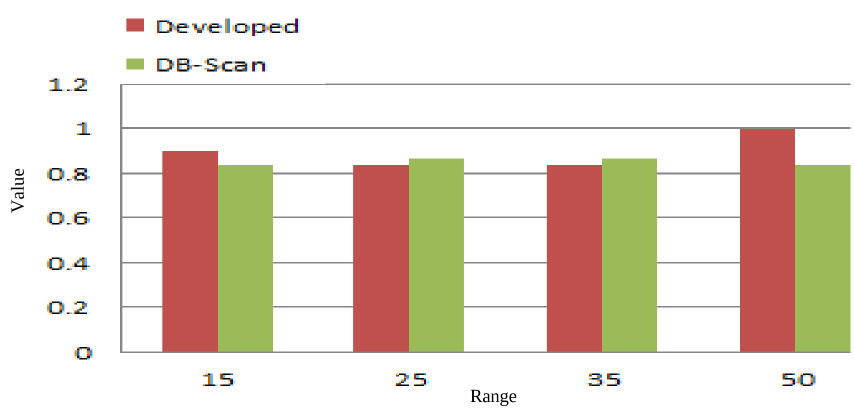

Fig. 11: Effects of different transmission ranges on the end-to-end delay values

and the network metrics in the two clustering algorithms. Figure 9 shows the effects of varying the nodes transmission ranges on theNRL in the two clustering algorithms ("DB-Scan and thedeveloped clustering").

Figure 9 shows that the average NRL values of the developed clustering approach are dominated the average NRL values of the DB-Scan at all the transmission ranges. Figure 10 show the shows the effects of varying the nodes transmission ranges on the PDFin the two clustering algorithms.

Figure 10 shows that the average PDF of the developed clustering are less than the average NRL values of the DB-Scan at all the transmission ranges.

Figure 11 shows the relationship between average End-to-end delay in the two algorithms and nodes transmission ranges.

Figure 11 shows that the average End-to-end delay valuesof the developed clustering are less than the average End-to-end valuesof the DB-Scan when the transmissionranges are (25 and 35) but when the transmission ranges (15 and 50) the average End-to-end valuesof the developed clustering are more than the average value of the "End-to-end delay"in the DB-Scan algorithm. Figure 12 shows the relationship between the average throughput in the two algorithms and transmission ranges.

Figure 12 shows that the average throughput of the developed clustering is less than its valuein DB-Scan when the transmission range is 25. The two values are the same when the transmissionrangeis 15. Whenthe transmission ranges are (35 and 50) the average throughput of the developed clustering are more than the average throughput of the DB-Scan.

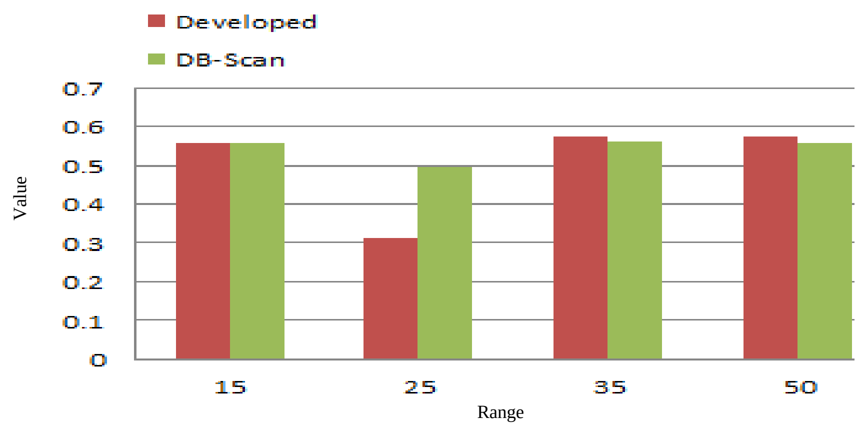

Fig. 12: Effects of different transmission ranges on throughput values

CONCLUSION

In this study the performance of a new clustering algorithm was suggested, simulated and implemented for suggested WSNs. The proposed approach was created, simulated and observed in order to test and evaluate their behaviors. Forming network clusters in a suitable manner is an effective way of improving the data transmission, network scalability and other behavior factors of the WSNs.The complex nature of the WSNs and several possible cluster configurations make searching for an optimal network structure is a still open challenge.

The simulation results have demonstrated the efficiency of the developed clustering approach. Compared with other approaches, this simulation study results demonstrated that the proposed algorithm improves the network performance to a certain level. Continuing in this approach will develop better applicable results.

REFERENCES

Alexandre, B., 2004. Multipath DSR protocol for ad hoc network. M.Sc Thesis, Institute of Electronic Systems, Aalborg University, Aalborg, Denmark.

Aliff, U.S., Z. Ishak, N.M. Din and M.Z. Jamaludin, 2006. Trace analyzer for NS-2. Proceeding of the 4th Student Conference on Research and Development (SCOReD), June 27-28, 2006, IEEE, Shah Alam, Malaysia, pp: 29-32.

Ducksoo, S., J. Lee, J. Kim and Jooseok, 2009. A2OMDV: An adaptive ad hoc on-demand multipath distance vector routing protocol using dynamic route switching. J. Eng. Sci. Technol., 4: 1-13.

Mamta, 2014. Various clustering techniques in wireless sensor network. Int. J. Comput. Appl. Technol. Res., 3: 381-384.

Priya V. and M. Chouhan, 2014. Survey on clustering techniques in wireless sensor network. Int. J. Comput. Sci. Inf. Technol. (IJCSIT.), 5: 6614-6619.

Saad, T.H., 2013. Designing a new MANET's environment using computer simulation. Int. J. Comput. Sci. Electr. Eng. (IJCSEE.), 1: 370-375.

Soni, T.M., 2012. An overview on clustering. IOSR J. Eng., 2: 719-725.