International Multilingual Journal of Science and Technology (IMJST)
ISSN: 2528-9810
Vol. 5 Issue 6, June - 2020

Application Of K-Means Clustering Algorithm For Selection Of Relay Nodes In Wireless Sensor Network

Ozuomba Simeon
Department of Electrical/Electronic and Computer Engineering,
University of Uyo, Akwa Ibom, Nigeria
simeonoz@yahoomail.com
simeonozuomba@uniuyo.edu.ng

***Abstract*— In this paper, the application of K-means clustering algorithm for selection of relay nodes in wireless sensor network (WSN) was presented. The case study WSN has 120 sensor nodes that are randomly distributed in a region of 1200 m by 1200 m. The base station (sink or sink node) was located at 600 m by 600 m which is the centre of the entire network coverage area. This is to ensure maximum signal transmission from all the nodes to the sink. The data for the simulation was generated using the Matlab random number generator function with normal distribution. The k-mean algorithm for the relay node selection was simulated in Matlab 2015a for two different hardware capacity threshold of 4.5 and 4.0. With hardware capacity threshold of 4.5 only 5 relay nodes were selected out of the 120 WSN sensor nodes. The cause of the small number of relay node is the high hardware capacity requirement for a sensor node to be considered as a relay node. This is confirmed from the number (26) of relay nodes selected when the threshold hardware capacity requirement was reduced to 4.0. Finally, the clustering of the WSN slave nodes with k-means algorithm around the 5 selected relay nodes with 4.0 threshold hardware capacity was also conducted. In all, the results show that the smaller the threshold of hardware capacity requirement for the relay nodes the more the number of relay nodes that are selected by the K-means algorithm. However, lowering the hardware capacity may lead to a relay node that cannot effectively handle the high demand on the lean hardware and network resources required for a relay node to deliver acceptable quality of service as relay node.**

Keywords*— K-Means Algorithm, Wireless Sensor Network, Relay Node, Clustering, Hardware Capacity, Clustering Algorithm.

I. INTRODUCTION

Today, Wireless Sensor Networks (WSNs) are increasingly applied in divers areas like precision agriculture, smart homes, and smart cities, etc [1,2,3,4,5]. WSN uses very small resource-limited sensors to capture environmental parameters and transmit the data to remote locations where the data can be utilized or stored [6,7,8,9,10,11]. Notably, the WSN sensors devices are mostly battery-powered which in most cases, the batteries have low storage capacities [12,13,14,15,16]. As such, effort is usually made to conserve the energy or manage the required transmission power for the sensor nodes. One way of minimizing the overall average transmission energy requirement in a network is by the use of clustering [17,18,19,20,21,22].
When a sensor device transmit signal, the amount of energy required depends on the distance between the sensor device and the destination remote device that will receive the data [23,24,25,26,27]. The longer the distance, the higher the power demand. Conversely, the shorter the distance between the sensor device and the receiving device, the smaller the transmission energy demand. Consequently, clustering is used along with device-to-device communication to reduce the transmission energy demand since the sensor nodes are enabled to transmit their data to the nearby relay nodes which requires less energy than transmitting directly to the sink (base station) which is far away from the sensor device. However, for clustering to be effective, the relay nodes (cluster) heads must be properly selected keeping in mind their capacity to handle such high transmission power and storage requirements [28,29,30,31,32,33]. As such, in this paper, the hardware capacity of the sensor nodes is included as key parameter used in the selection of the relay node. Specifically, K-means algorithm was used to do the relay node selection and for clustering of the WSN slave nodes to the relay node. The simulation was implemented in MATLAB software version 2015a.

II. METHODOLOGY

In this study, K-Means clustering algorithm was used for relay node selection in wireless sensor network (WSN). A WSN with 120 sensor nodes that are randomly distributed in a region of 1200 m by 1200 m was considered in the study. The base station (sink or sink node) was located at 600 m by 600 m which is the centre of the entire network coverage area. This is to ensure maximum signal transmission from all the nodes to the sink. The x and y location coordinates of each of the 120 WSN nodes were randomly generated using Matlab random number generator function with coordinate value in the range of 0 to 1200 . The Matlab software implementation of the

random distribution function ($randnd$) for the x and y set of coordinate data is given as follows;

$$randnd = lb + (ub - lb) \times rand(120,1) \quad (1)$$

where ub indicates upper value which is 1200m and lb indicates the lower value which is 0m .

In the same way, hardware capacity of WSN sensor devices (denoted as hc) was randomly generated with the values ranging from 0 to 5 which is defined using the following composite function;

$$\left.\begin{array}{ll} 4 < hc \leq 5; & \text{very strong capacity.} \\ 3 < hc \leq 4; & \text{strong capacity} \\ 2 < hc \leq 3; & \text{fairly strong hardware capability} \\ 1 < hc \leq 2; & \text{poor capability} \\ hc \leq 1; & \text{very poor } capability \end{array}\right\} \quad (2)$$

The Matlab software implementation of the random distribution function ($randnd$) for the hardware capacity of WSN sensor devices is given as follows;

$$randnd = lhc + (uhc - lhc) \times rand(120,1) \quad (3)$$

The generated x-y location coordinates of the WSN sensor nodes and their hardware capacity shown in Table 1. The plot of the x,y location or position coordinates of the WSN nodes and the base station is shown in Figure 1.

Table 1; The generated x-y location coordinates of the WSN sensor nodes and their hardware capacity

Device number	Hardware Capacity	x-coordinate (m)	y-coordinate (m)	Device number	Hardware Capacity	x-coordinate (m)	y-coordinate (m)
1	4.8763	131.6786	105.5857	61	3.6374	675.3802	73.6106
2	4.613	1187.659	324.0829	62	1.3891	365.8563	842.0209
3	0.7215	5.7218	988.9993	63	2.3376	919.4559	52.3885
4	3.2581	1009.104	1146.812	64	3.3728	233.3031	88.2127
5	4.375	956.7638	36.0773	65	4.6234	847.9474	644.0708
6	0.5243	1072.525	901.728	66	2.3674	226.5741	119.4347
7	4.3498	987.8263	566.5329	67	4.9384	454.959	1010.153
8	4.8174	320.8522	292.969	68	0.6783	772.4364	1009.413
9	2.8746	493.6045	714.2932	69	0.6374	963.3282	891.9843
10	1.3562	104.2481	603.2749	70	1.1283	100.1644	185.0374
11	4.7162	1124.382	675.1333	71	4.9384	1147.532	814.4052
12	2.4452	532.6615	1189.1	72	1.4382	957.7619	640.2948
13	2.8174	325.7086	285.9416	73	4.2189	601.0319	1201.344
14	4.7984	224.5266	643.4329	74	1.6172	538.1456	801.297
15	4.9904	179.6974	603.6351	75	4.9283	551.6344	988.1556
16	0.2096	167.9988	770.5123	76	1.2384	378.2442	560.2978
17	2.321	1073.349	838.5123	77	0.6231	627.8451	533.8653
18	3.2736	678.8991	453.6626	78	4.7123	630.6437	1019.002
19	3.4283	1053.182	46.5958	79	3.1212	1009.505	103.0546
20	0.1328	768.043	1092.898	80	4.7234	981.3648	164.4285
21	4.3726	715.7543	488.3365	81	1.9821	795.5249	214.0832
22	4.9847	178.9772	1219.864	82	4.2634	467.4601	482.6879
23	3.3429	496.1047	121.8732	83	4.8345	1002.046	1026.492
24	3.8126	93.7945	323.3318	84	2.7283	657.8684	991.9008
25	4.8371	433.3109	1127.572	85	4.8394	433.0331	74.6621
26	4.3928	633.6995	983.0316	86	1.8273	1159.365	492.9563
27	3.3987	296.2203	414.0602	87	3.8293	1081.477	650.522
28	1.8219	152.2566	839.2429	88	3.8273	679.2695	514.6146

29	3.3817	227.0679	168.5956	89	1.8123	768.5574	811.0098
30	0.9374	296.2615	890.4821	90	2.7162	724.8086	775.3482
31	3.2091	515.1908	131.8124	91	0.7283	256.4945	360.506
32	0.2784	61.307	807.1823	92	0.8172	371.9474	532.9496
33	1.3624	1114.607	610.148	93	2.7834	581.4417	19.1221
34	0.5091	1166.464	961.8775	94	3.7734	284.5835	1215.028
35	0.8457	606.0633	882.8374	95	4.5526	1042.482	206.3973
36	3.6459	416.972	1099.997	96	0.8234	240.4745	131.1436
37	4.1734	604.0672	1115.842	97	2.3374	278.9451	459.8051
38	0.7234	137.2964	244.23	98	2.1728	210.7702	244.6107
39	3.0834	455.9056	862.7224	99	0.1772	281.0955	604.6125
40	4.7834	455.9056	862.7224	100	1.2273	537.9501	419.1636
41	2.1263	963.3591	37.7082	101	0.8394	384.1089	1175.004
42	1.8475	481.2062	918.6945	102	4.1283	1140.124	1136.317
43	3.1734	298.4119	617.3709	103	1.2263	531.1696	65.0388
44	3.3094	498.6975	592.5538	104	2.4884	228.1894	911.0189
45	0.8219	119.0952	1117.077	105	0.5394	1117.283	332.273
46	2.0938	162.9469	752.9902	106	3.2832	1209.678	522.0639
47	2.791	1163.172	762.6207	107	1.8834	541.8599	676.4503
48	3.7263	1180.56	1061.105	108	3.4162	137.1935	1163.995
49	3.9043	710.1982	994.5245	109	3.3398	318.6298	515.7773
50	3.2093	73.8081	712.0708	110	4.2384	504.6343	1213.793
51	1.7623	289.8823	225.8538	111	2.0293	734.5111	372.2046
52	4.9283	801.4514	1208.34	112	0.7092	323.7434	865.6341
53	3.7734	1013.909	1094.544	113	1.3235	744.3165	822.7187
54	0.9123	19.0182	35.4034	114	4.8384	878.125	665.6469
55	0.8344	53.1211	604.8697	115	0.3728	273.7903	861.9404
56	2.9112	208.6506	207.3336	116	4.2234	144.972	822.9451
57	3.9028	436.0375	296.2409	117	2.8374	366.2987	219.9377
58	1.6374	903.4462	879.9461	118	2.3849	393.5851	158.0596
59	2.1892	799.764	617.9265	119	4.8849	523.7101	1233.548
60	1.0231	556.7481	581.6475	120	2.8374	627.0425	211.2846

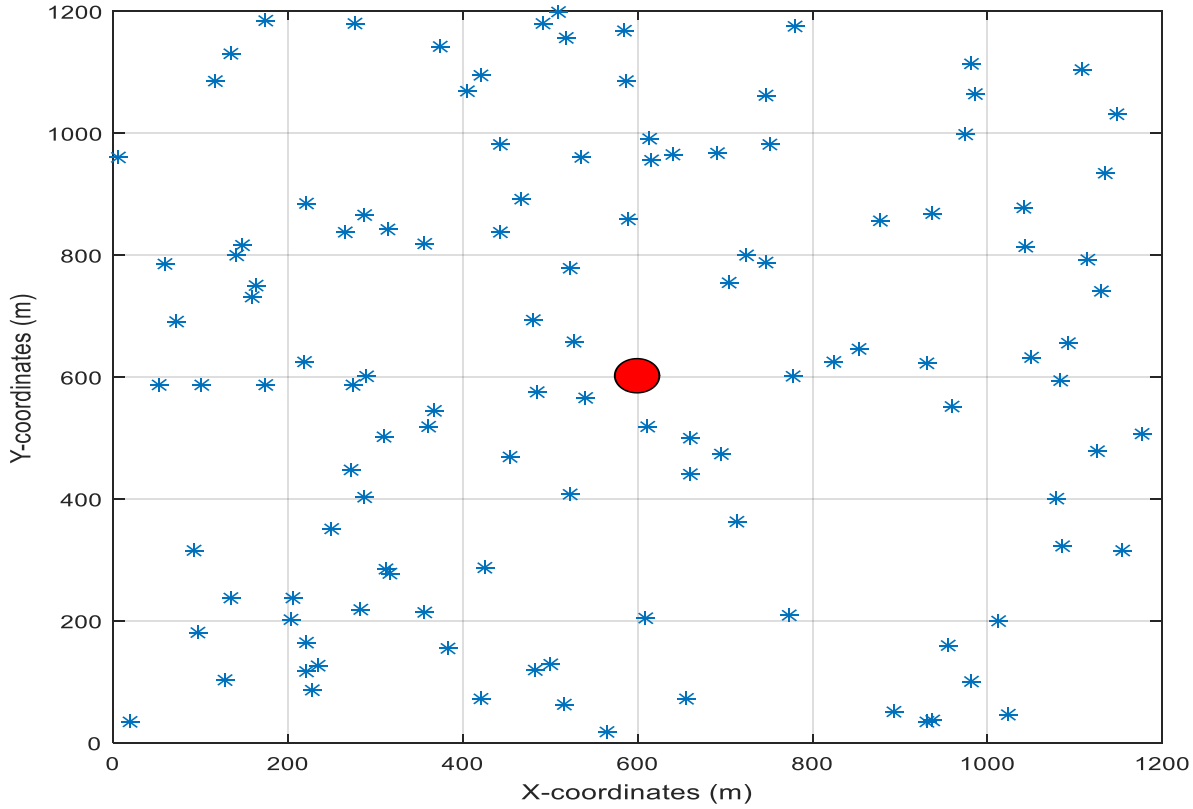

Figure 1; The plot of the x,y location or position coordinates of the WSN nodes and the base station.

III. RESULTS AND DISCUSSION

MATLAB software version 2015a was used to simulate the relay node selection based on K-means algorithm as well as the clustering ot the remaining WSN slave nodes to the selected relay nodes. The result of the relay node selection is shown in Table 2, which shows the x-y location coordinates of the selected relay nodes and their corresponding hardware capacities. Also, Figure 2 shows the graphical plot of the x-y location coordinates of the selected relay nodes. Notable, the selection of relay nodes in this paper was based on the nodes with hardware capacity greater than 4.5 and above. As such, the results in Table 2 show the relay nodes with hardware capacity that are above the threshold of 4.5. Table 2 also shows that only 5 relay nodes were selected out of the 120 WSN sensor nodes. The cause of the small number of relay node is the high hardware capacity requirement for a relay node. This is confirmed from the number (26) of relay nodes selected when the threshold hardware capacity requirement was reduced to 4.0 , as shown in Table 3 and Figure 3. Finally, Figure 4 shows the clustering of the WSN slave nodes with k-means algorithm around the 5 selected relay nodes with 4.0 threshold hardware capacity. In all, the results show that the smaller the threshold of hardware capacity requirement for the relay nodes the more the number of relay nodes that are selected by the K-means algorithm. However, lowering the hardware capacity may lead to a relay node that cannot effectively handle the high demand on the lean hardware and network resources required for a relay node to deliver acceptable quality of service as relay node.

Table 2; The selected 5 relay nodes , their x-y location coordinates of the selected relay nodes and their corresponding hardware capacities.

CH number (device number)	x-coordinates (m)	y-coordinates (m)	Hardware Capacity
1(1)	1154.3	314.98	4.529
2(2)	311.84	284.74	4.7875
6(11)	4.7162	1124.38	675.133
8(15)	4.9904	179.697	603.635
14(37)	4.1734	604.067	1115.84

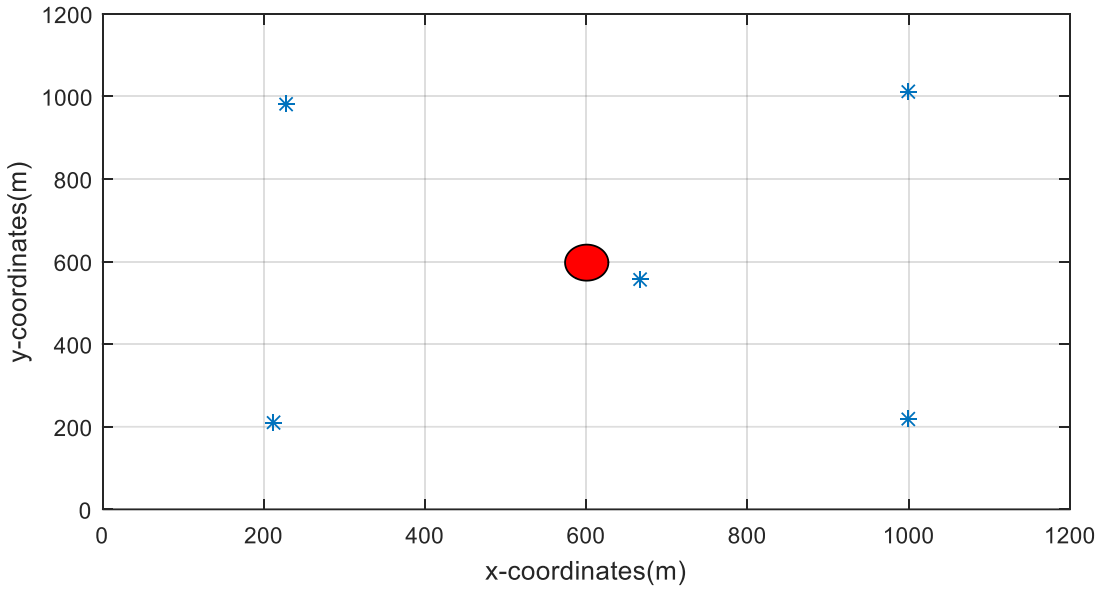

Figure 2; The graphical plot of the x-y location coordinates of the 5 selected relay nodes with thresh hold hardware capacity of 4.5.

Table 3; The selected 26 relay nodes , their x-y location coordinates of the selected relay nodes and their corresponding hardware capacities with thresh hold hardware capacity of 4.0.

CH number (device number)	x-coordinates (m)	y-coordinates (m)	Hardware Capacity
1(1)	4.8763	131.679	105.586
2(2)	4.613	1187.66	324.083
3(5)	4.375	956.764	36.0773
4(7)	4.3498	987.826	566.533
5(8)	4.8174	320.852	292.969
6(11)	4.7162	1124.38	675.133
7(14)	4.7984	224.527	643.433
8(15)	4.9904	179.697	603.635
9(21)	4.3726	715.754	488.337
10(22)	4.9847	178.977	1219.86
11(25)	4.8371	433.311	1127.57
12(26)	4.3928	633.7	983.032
13(37)	4.1734	604.067	1115.84
14(40)	4.7834	838.49	4.7511
15(52)	4.9283	801.451	1208.34
16(65)	4.6234	847.947	644.071
17(67)	4.9384	116.08	4.7965
18(75)	4.9283	551.634	988.156
19(78)	4.7123	630.644	1019
20(80)	4.7234	981.365	164.429
21(83)	4.8345	1002.05	1026.49
22(95)	4.5526	1042.48	206.397
23(102)	4.1283	1140.12	1136.32
24(110)	4.2384	504.634	1213.79
25(114)	4.8384	878.125	665.647
26(119)	4.8849	523.71	1233.55

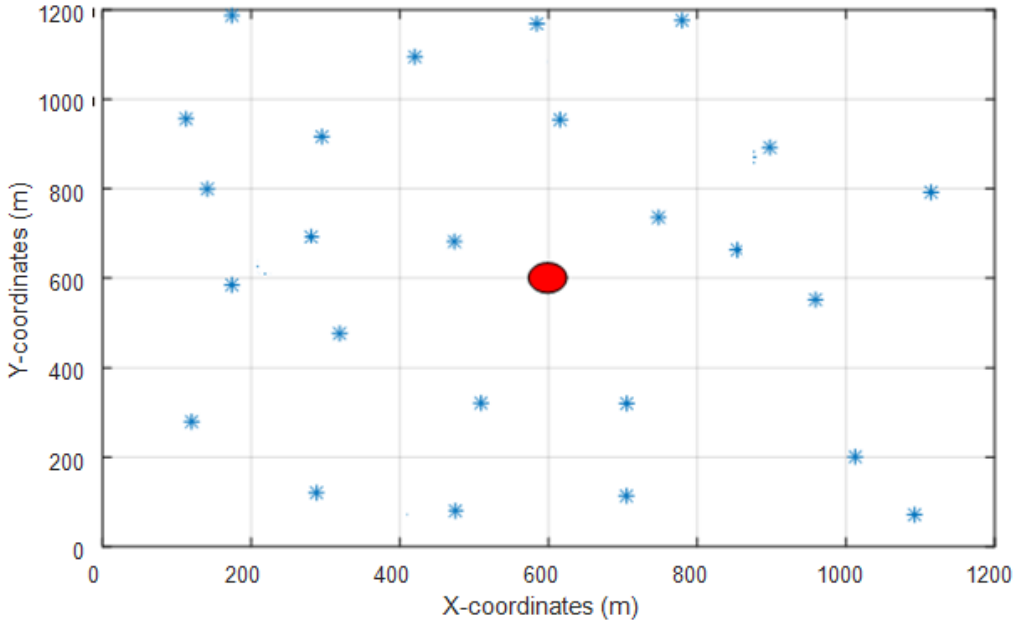

Figure 3; The graphical plot of the x-y location coordinates of the 26 selected relay nodes with thresh hold hardware capacity of 4.0.

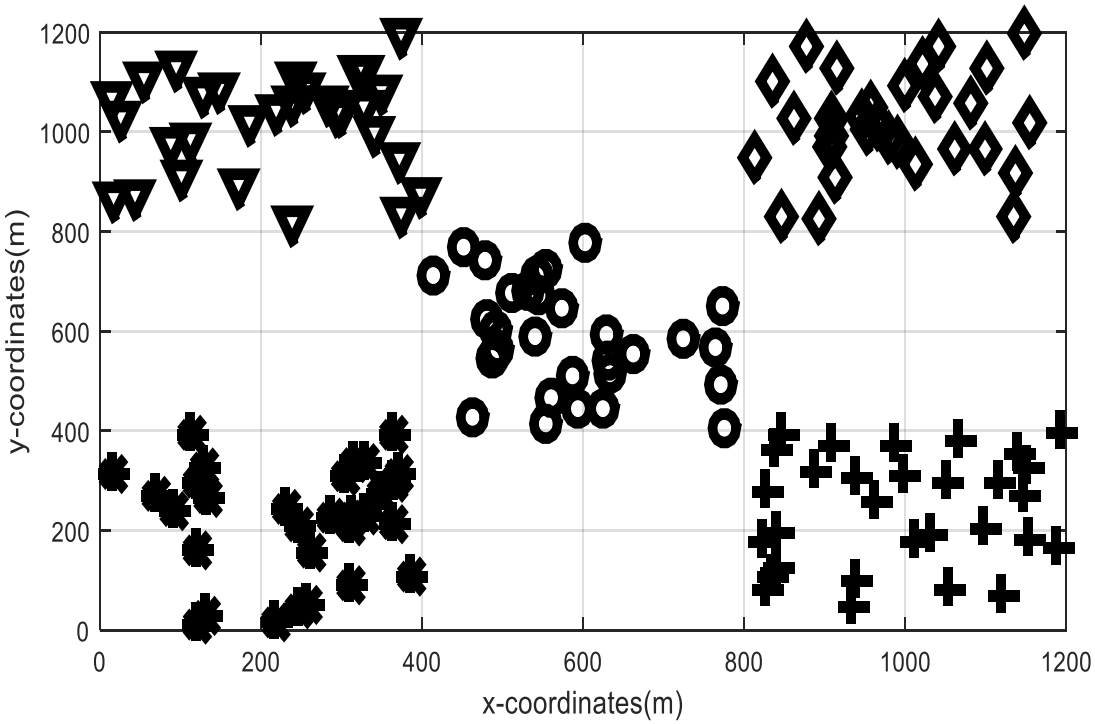

Figure 4; Clustering of WSN with k-means algorithm around the 5 selected relay nodes with 4.0 threshold hardware capacity

IV. CONCLUSION

The use of K-means algorithm in the selection of relay nodes in wireless sensor network is presented. The K-means algorithm was also used to cluster the slave sensor nodes to the relay nodes. The relay nodes selection is based on the hardware capacity of the sensor devices and the distance of the nodes from the sink node which is the base station. The data for the simulation was generated using the Matlab random number generator function with normal distribution. The simulation for the relay node selection was conducted with two different hardware capacities of 4.5 and 4.0. The results shows that the higher the hardware capacity threshold, the smaller the number of relay nodes the K-means algorithm was able to select.

REFERENCES

1. Hryniewicz, T. (2016). UGC Sponsored Two Day National Conference on Internet of Things 18th and 19th February 2016. *World Scientific News*, (41), 1-304.
2. Foley, J. G. (2014). *Sensor networks and their applications: Investigating the role of sensor web enablement* (Doctoral dissertation, UCL (University College London)).

3. Farooq, M. S., Riaz, S., Abid, A., Umer, T., & Zikria, Y. B. (2020). Role of IoT Technology in Agriculture: A Systematic Literature Review. *Electronics*, *9*(2), 319.
4. Glória, A., Dionísio, C., Simões, G., Cardoso, J., & Sebastião, P. (2020). Water Management for Sustainable Irrigation Systems Using Internet-of-Things. *Sensors*, *20*(5), 1402.
5. González-Briones, A., Castellanos-Garzón, J. A., Mezquita Martín, Y., Prieto, J., & Corchado, J. M. (2018). A framework for knowledge discovery from wireless sensor networks in rural environments: a crop irrigation systems case study. *Wireless Communications and Mobile Computing*, *2018*.
6. Bader, S. (2011). *Enabling autonomous envionmental measurement systems with low-power wireless sensor networks* (Doctoral dissertation, Mittuniversitetet).
7. Othman, M. F., & Shazali, K. (2012). Wireless sensor network applications: A study in environment monitoring system. *Procedia Engineering*, *41*, 1204-1210.
8. Djedouboum, A. C., Abba Ari, A. A., Gueroui, A. M., Mohamadou, A., & Aliouat, Z. (2018). Big data collection in large-scale wireless sensor networks. *Sensors*, *18*(12), 4474.
9. Sen, J. (2010). A survey on wireless sensor network security. *arXiv preprint arXiv:1011.1529*.
10. Zennaro, M., Pehrson, B., & Bagula, A. (2008, October). Wireless Sensor Networks: a great opportunity for researchers in Developing Countries. In *Proceedings of WCITD2008 Conference, Pretoria, South Africa* (Vol. 67).
11. Al-Karaki, J. N., & Kamal, A. E. (2004). Routing techniques in wireless sensor networks: a survey. *IEEE wireless communications*, *11*(6), 6-28.
12. Wenqi, G. U. O., & Healy, W. M. (2014). Power Supply Issues in Battery Reliant Wireless Sensor Networks: A. *International Journal of Intelligent Control and Systems*, *19*(1), 15-23.
13. Rodrigues, L. M., Montez, C., Budke, G., Vasques, F., & Portugal, P. (2017). Estimating the lifetime of wireless sensor network nodes through the use of embedded analytical battery models. *Journal of Sensor and Actuator Networks*, *6*(2), 8.
14. Engmann, F., Katsriku, F. A., Abdulai, J. D., Adu-Manu, K. S., & Banaseka, F. K. (2018). Prolonging the lifetime of wireless sensor networks: A review of current techniques. *Wireless Communications and Mobile Computing*, *2018*.
15. Healy, M., Newe, T., & Lewis, E. (2008, October). Wireless sensor node hardware: A review. In *SENSORS, 2008 IEEE* (pp. 621-624). IEEE.
16. Kamal, A., & Patel, P. Dynamic Slot Allocation for Improving Traffic in Wireless Sensor Network: A Survey.
17. Eleburuike, I., & Soetan, S. A. (2010). An Energy Efficient Protocol Based on Hierarchical Routing Technique for WSN.
18. Li, W., Martins, P., & Shen, L. (2012). Determination method of optimal number of clusters for clustered wireless sensor networks. *Wireless Communications and Mobile Computing*, *12*(2), 158-168.
19. Tinker, M. S., & Chinara, S. (2015, May). Energy conservation clustering in wireless sensor networks for increased life time. In *2015 Second International Conference on Advances in Computing and Communication Engineering* (pp. 7-10). IEEE.
20. Elshrkawey, M., Elsherif, S. M., & Wahed, M. E. (2018). An enhancement approach for reducing the energy consumption in wireless sensor networks. *Journal of King Saud University-Computer and Information Sciences*, *30*(2), 259-267.
21. Bandyopadhyay, S., & Coyle, E. J. (2003, March). An energy efficient hierarchical clustering algorithm for wireless sensor networks. In *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No. 03CH37428)* (Vol. 3, pp. 1713-1723). IEEE.
22. Liu, X., & Wu, J. (2019). A method for energy balance and data transmission optimal routing in wireless sensor networks. *Sensors*, *19*(13), 3017.
23. Pour, N. K. (2016). Energy efficiency in wireless sensor networks. *arXiv preprint arXiv:1605.02393*.
24. Ghenghea, M., Darie, A., & Bogdan, I. (2010). Wireless Sensor Network Interface for a Residential Monitoring System. *IFAC Proceedings Volumes*, *43*(23), 203-207.
25. Ketshabetswe, L. K., Zungeru, A. M., Mangwala, M., Chuma, J. M., & Sigweni, B. (2019). Communication protocols for wireless sensor networks: A survey and comparison. *Heliyon*, *5*(5), e01591.
26. Alves, A. J. R., Manera, L. T., & Campos, M. V. (2019). Low-cost wireless sensor network applied to real-time monitoring and control of water consumption in residences. *Revista Ambiente & Água*, *14*(6).
27. Grindvoll, H., Vermesan, O., Crosbie, T., Bahr, R., Dawood, N., & Revel, G. M. (2012). A wireless sensor network for intelligent building energy management based on multi communication standards-A case study. *Journal of Information technology in construction*, *17*, 43-61.
28. Rahman, T., Ullah, I., Rehman, A. U., & Naqvi, R. A. (2020). Clustering Schemes in MANETs: Performance Evaluation, Open Challenges, and Proposed Solutions. *IEEE Access*, *8*, 25135-25158.
29. Han, T., Bozorgi, S. M., Orang, A. V., Hosseinabadi, A. A. R., Sangaiah, A. K., & Chen, M. Y. (2019). A hybrid unequal clustering based on density with energy conservation in wireless nodes. *Sustainability*, *11*(3), 746.

30. Tripathy, A. K., & Chinara, S. (2012). Comparison of residual energy-based clustering algorithms for wireless sensor network. *International Scholarly Research Notices*, *2012*.
31. Ali, A. E. (2019). *Reliable and energy efficient mechanisms for wireless sensor networks* (Doctoral dissertation, Edinburgh Napier University, UK).
32. Kumar, A. (2017). *Energy efficient clustering algorithm for wireless sensor network* (Doctoral dissertation, Lovely Professional University).
33. Montiel, E. R., Rivero-Angeles, M. E., Rubino, G., Molina-Lozano, H., Menchaca-Mendez, R., & Menchaca-Mendez, R. (2017). Performance analysis of cluster formation in wireless sensor networks. *Sensors*, *17*(12), 2902.