

# Evaluation Of Lora-Transceiver Transmission Phase Energy Consumption For Clustered Sensor Network

Ezuruike Okafor S. F. <sup>1</sup>

Department of Electrical/Electronic Engineering  
Imo State Polytechnic,  
Omuma, Imo State, Nigeria

Maduka N.C. <sup>2</sup>

Department of Physics  
Federal University Gusau,  
Zamfara State Nigeria  
madukanosike@fugusau.edu.ng

Agbaji Adai Samue<sup>3</sup>

Department of Computer Engineering  
University of Uyo, Uyo  
Uyo, Akwa Ibom State, Nigeria  
agajisam@yahoo.com

**Abstract**— In this paper, evaluation of LoRa-transceiver transmission phase energy consumption for clustered sensor network is presented. The transmitter phase of sensor node is the stage in the communication process at which the data is transmitted from the LoRa transceiver to the receiver which in a clustered network is the gateway or the cluster head. The transmitter power is derived from link power budget expression and the propagation loss,  $P_{Loss}$  is based on a modified free space model with path loss exponent  $n$ . The clustering was done using gap statistics for optimal number of cluster determination and then clustering using K-means algorithm with early centroid determination based-on mean. The simulation was conducted with randomly distributed 2000 sensor nodes in an area of 3.5 km x 3.5 km. The Gap statistics method gave optimal number of clusters as 5 and the K-means algorithm was then used to cluster the sensor nodes into 5 clusters. The results show that the overall Average Euclidian Distance (AED) of the sensor nodes in the network is 60.48 m, Cluster 1 has the highest AED with a value of 652.27 m and the sensor node with the highest Maximum Euclidian Distance (SMxED) with a value of 1205.47 m. On the other hand, Cluster 5 has the lowest mean AED with a value of 568.67 m. Cluster 1 has the sensor node with the highest Transmitter Energy Consumption (TEC) with a value of 8738.5 mw or 39.4 dBm for SF of 12 and a value of 11225.8 mw or 40.5 dBm for SF of 7. Also, the overall Average Transmitter Energy Consumption (ATEC) of the sensor nodes in the network is 1447.8 mw or 31.6 dBm for SF of 12. In all, although the spreading factor, SF of 7 required higher transmitter power than the SF of 12, the

energy consumption of SF of 12 is higher than that of the SF of 7 because of the higher transmission time required by the SF of 12.

**Keywords**— Gap Statistics Technique, Lora-Transceiver, Transmission Phase Energy Consumption, Clustering Algorithm, Sensor Network, Classical K-means Algorithm

## 1. INTRODUCTION

In recent years, there has been drastic rise in the adoption of smart technologies across the globe [1,2,3]. The smart technologies relies on robust wireless sensor technologies and artificial intelligence programs [4,5,6]. While many transceiver technologies exists for the wireless sensors, the Long Range low power (LoRa) transceiver technology has stood out as the most popular due to its numerous salient qualities [7,8].

Typically, LoRa has flexile adaptive data rate technology that enable it to guarantee different long distance transmissions at different data rates and with different power consumption rates [9,10]. Several studies have compared the LoRa range and energy consumption combinations and the conclusions have been that LoRa is efficient in the transmission energies in covering the required transmission range. As such, LoRa technology has been deployed in direct earth - to - satellite transmission [11,12,13]. Accordingly, this paper examines the energy consumption in LoRa transceiver during the data transmission phase, particularly when the LoRa transceiver is used in a clustered sensor network. The study is meant to understand the distribution of the sensor node energy consumption in each cluster and to identify the critical

cluster in the network. Such study will enable the network designer to select the appropriate LoRa transceiver parameters combinations that will guarantee the specified quality of service for a given minimum network lifespan when the sensors are powered using battery.

## 2. METHODOLOGY

The transmitter phase of sensor node is the stage in the communication process at which the data is transmitted from the LoRa transceiver to the receiver which in a clustered network is the gateway or the cluster head. The energy consumed in the transmitter phase (denoted as  $E_{tx}$ ) is given in terms of the transmitter power ( $P_{tx}$ ) and the transmission time ( $t_{tx}$ ).

$$E_{tx} = (P_{tx})(t_{tx}) \quad (1)$$

### 2.1 Determination of the LoRa transceiver transmission time, $t_{tx}$

For LoRa transceiver, the transmission time is defined as the packet time on air, which is expressed analytically as follows [14,15];

$$t_{tx} = (n_{PL} + n_{PR} + 4.25)T_s \quad (2)$$

$$n_{PL} = 8 + \max \left( \left( \text{ceil} \left[ \frac{8PL-4SF+28+16 \text{ CRC}-20H}{4(SF-2DE)} \right] (CR + 4) \right), 0 \right) T_s \quad (3)$$

$$T_s = \frac{1}{R_s} = \frac{2^{SF}}{BW} \quad (4)$$

$n_{PR}$  indicates the packet preamble size in bytes ; SF indicates the LoRa spreading factor; BW indicates the LoRa bandwidth parameter which has different options as 125 KHz, 250 KHz or 500 KHz; PL indicates the payload size in bytes; H indicates header flag where  $H = 0$  shows that H is enabled and  $H = 1$  shows that h is in disabled state; DE indicates low data rate optimization where  $DE = 1$  shows that DE is enabled and  $DE = 0$  is for DE disabled state; CR indicates the forward error correction bit known as the coding rate which can take any 4 different values as CR 1, 2, 3, or 4 and CRC value is set at 1 for uplink and it is set at 0 for down link.

### 2.2 Determination of the LoRa transceiver transmitter power, $P$

The transmitter power is derived from link power budget expression with  $L_m$  as the required link margin,  $S_{LoRa}$  as the sensitivity of the LoRa transceiver,  $P_{Loss}$  as the propagation loss while  $G_{tx}$  and  $G_{rx}$  are the antenna gain for the transmitter and receiver respectively. Then [16,17];

$$P_{tx} = LM + S_{LoRa} - (G_{tx} + G_{rx}) + P_{Loss} \quad (5)$$

### 2.3 Determination of the propagation loss for the LoRa transceiver

The propagation loss,  $P_{Loss}$  is based on a modified free space model with path loss exponent  $n$  which is expressed as;

$$P_{Loss} = 32.45 + 10n \text{ Log}(f) + 10 n \text{ Log}(d) \quad (6)$$

Where the signal frequency ( $f$ ) is expressed in MHz while the transmission path length ( $d$ ) is expressed in km.

### 2.4 Determination of the propagation loss for the LoRa transceiver

Since the transmitter power is in dBm, the transmitter energy in milliwatt is given as;

$$E_{tx}(mW) = (t_{tx}) \left( 10^{\left( \frac{P_{tx}}{10} \right)} \right) \quad (7)$$

$$E_{tx}(dBm) = 10 \text{ Log} (E_{tx}(mW)) \quad (8)$$

### 2.5 Determination of the transmission path length for the LoRa transceiver based on sensor node clustering

The value of the transmission path length,  $d$  is obtained from the Euclidian distance computed from the clustered sensor nodes. The sensor nodes which are distributed randomly within the network area is clustered using the K-means algorithm with early centroid determination based-on mean and the procedure for clustering approach is given in Algorithm 1 [18].

Also, in order to determine the number of clusters,  $k$  used in the clustering algorithm 1, the Gap Statistics method (shown in flow diagram of Figure 1) is used to determine the optimal value of  $k$ . After the clustering, the various Euclidian distance parameters are computed. Consider each of the  $k$  clusters with the coordinates of the centroid of cluster  $j$  as  $(Cx_j, Cy_j)$  and the coordinates of the cluster member,  $i$  in cluster  $j$  as  $(x_{j,i}, y_{j,i})$  where  $j = 1, 2, 3, \dots, k$  and  $i = 1, 2, 3, \dots, n_j$ , where  $n_j$  is the number of sensor nodes in cluster  $j$ . The Euclidian distance,  $d_{j,i}$  of  $x_{j,i}, y_{j,i}$  from centroid  $Cx_j, Cy_j$  is determined as;

$$d_{j,i} = \sqrt{(Cx_j, Cy_j)^2 + (x_{j,i}, y_{j,i})^2} \quad (9)$$

Let  $d_j$  denote the mean Euclidian distance of cluster members in cluster  $j$ , then;

$$d_j = \left( \frac{1}{n_j} \right) \left( \sum_{i=1}^{n_j} (d_{j,i}) \right) \quad (10)$$

Let  $d_j$  denote the mean Euclidian distance of cluster members in cluster  $j$ , then;

Let  $d_{avg}$  denote the mean Euclidian distance of all the  $k$  clusters in the network, then;

$$d_{avg} = \left( \frac{1}{k} \right) \left( \sum_{j=1}^k (d_j) \right) \quad (11)$$

Let  $d_{max(j,i)}$  denote the maximum Euclidian distance of cluster member from its centroid in the network, then;

$$d_{max(j,i)} = \text{maximum}(d_{j,i}) \text{ for } j = 1, 2, 3, \dots, k \text{ and } i = 1, 2, 3, \dots, n_j \quad (12)$$

Let  $d_{max(j)}$  denote the maximum mean Euclidian distance of cluster in the network, then;

$$d_{max(j)} = \text{maximum}(d_j) \text{ for } j = 1, 2, 3, \dots, k \text{ a} \quad (13)$$

Let  $d_{min(j)}$  denote the minimum mean Euclidian distance of cluster in the network, then;

$$d_{\min(j)} = \text{minimum}(d_j) \text{ for } j = 1, 2, 3, \dots, k \text{ a} \quad (14)$$

**Algorithm 1:**

- Step 1. Input "The number of clusters", K
- Step 2. Input "The number of data items available in the dataset" N //
- Step 3. Input "The N number points,  $d_i$  where  $d_i$  for  $i = 1, 2, 3, \dots, N$  |
- Step 4. Group the N data items,  $d_i$  into K clusters
- Step 5. For each of the K clusters determine the initial centroid  $c_j$  by computing the mean of the data items,  $d_i$  in cluster j where  $1 \leq j \leq K$
- Step 6. Compute the Euclidean distance between each of the centroids and each of the data point and assign each data point to the cluster belonging to the centroid that it has the smallest Euclidean distance value
- Step 7. Compute the centroid again for the k clusters
- Step 8. Repeat Step 6 and Step 7 until none of the centroid values changes again

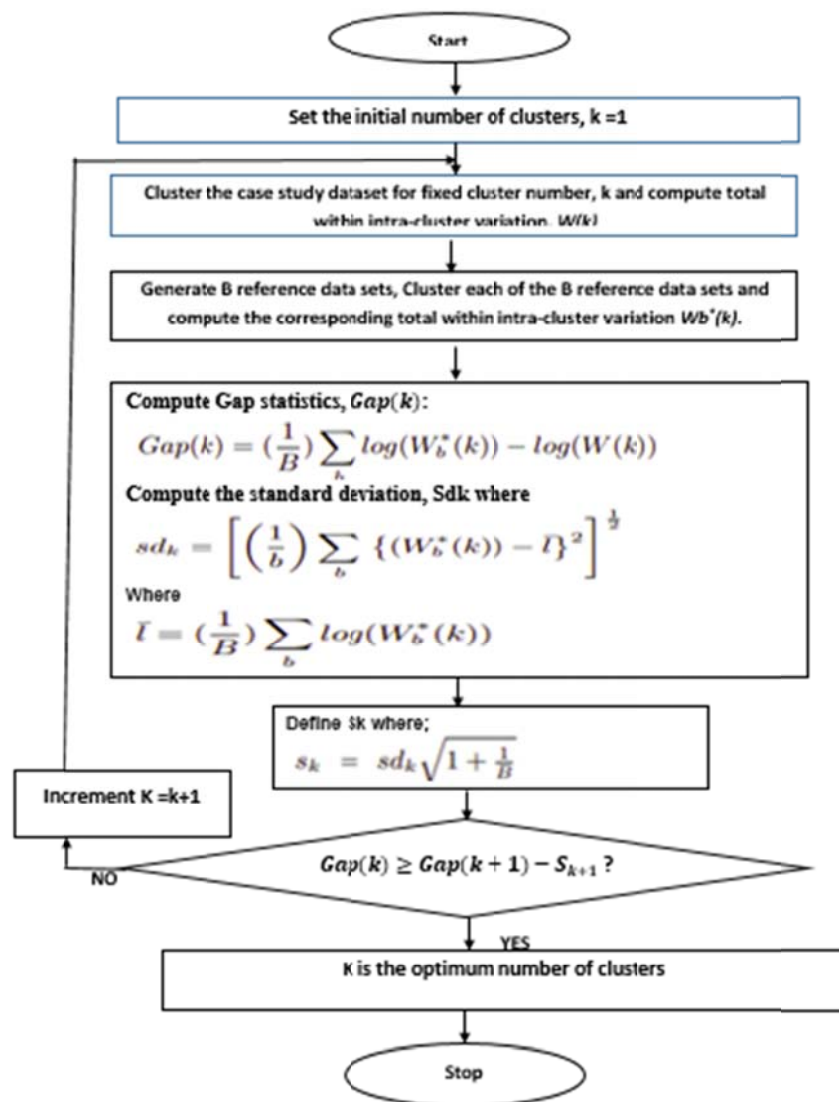


Figure 1 Gap Statistics flow diagram for optimal number of cluster determination

Each of the following Euclidian distance parameters,  $d_{max(j,i)}$ ,  $d_j$ ,  $d_{avg}$ ,  $d_{max(j,i)}$ ,  $d_{max(j)}$  and  $d_{min(j)}$  where used to determine the path loss and hence the energy consumption by the LoRa transceiver during the packet transmission phase. The simulation program was written in Visula Basic for Application (VBA)

**3. RESULTS AND DISCUSSIONS**

The simulation was conducted with randomly distributed 2000 sensor nodes in an area of 3.5 km x 3.5 km (shown in

Figure 1). The Gap statistics method gave optimal number of clusters as 5 and the K-means algorithm was then used to cluster the sensor nodes into 5 clusters 9as shown in Figure 2) while the visualization of the clustering algorithm convergence is shown in Figure 3. The LoRa receiver sensitivity at bandwidth of 125 KHz and the corresponding transmission time for the various spreading factors (SF) and with packet size of 51 bytes are shown in Table 2.

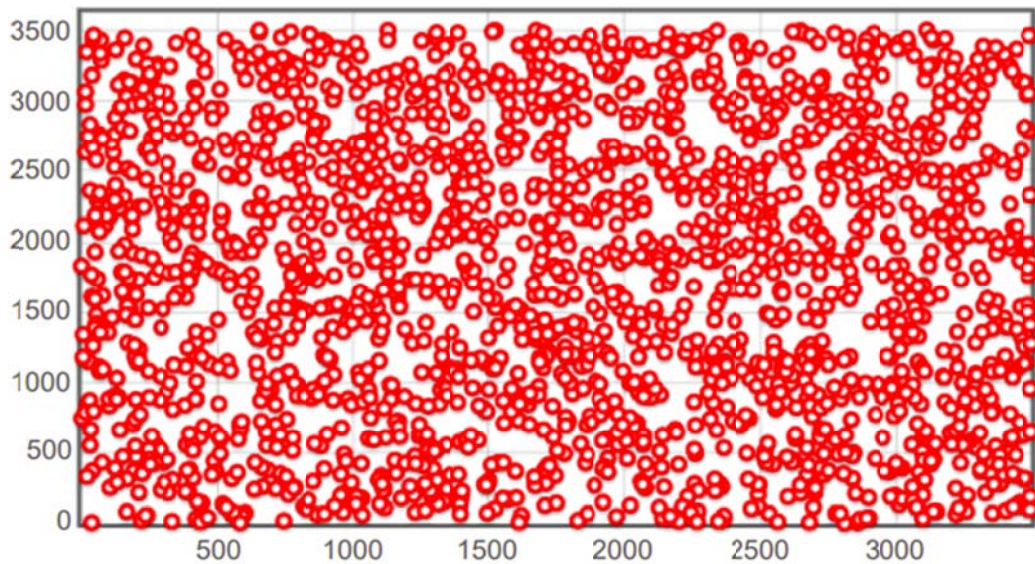


Figure 1 The visualization of the randomly distributed 2000 sensor nodes in an area of 3.5 km x 3.5 km

Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5

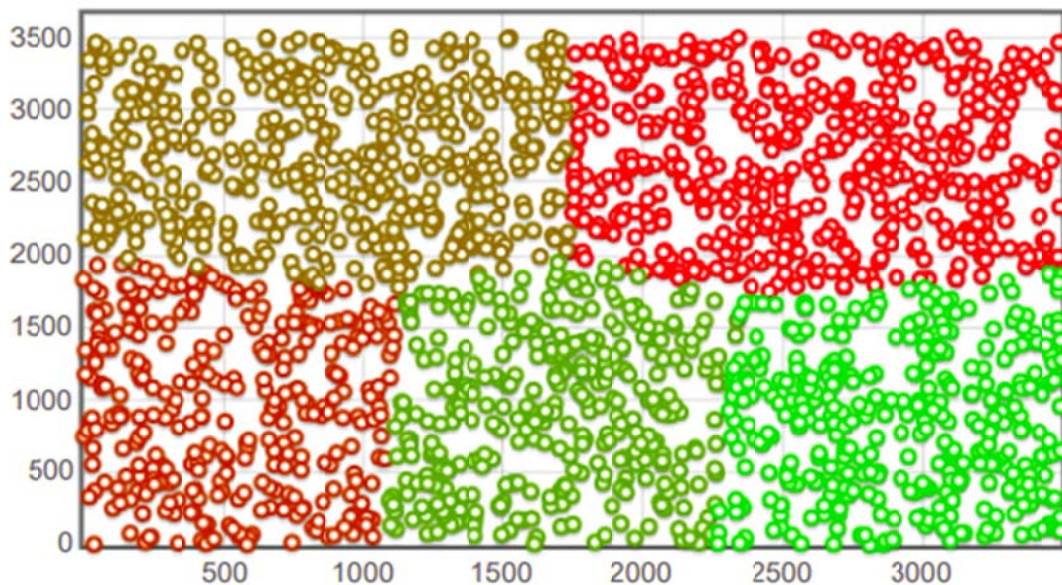
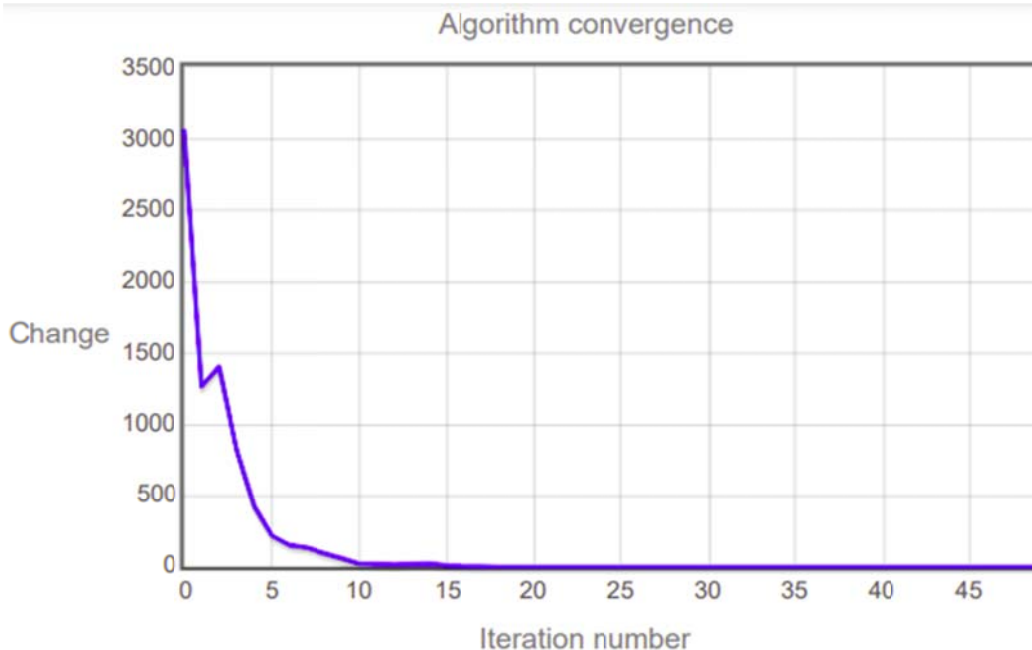


Figure 2 The visualization of the 2000 sensor nodes clustered into 5 clusters in an area of 3.5 km x 3.5 km



**Figure 3 The visualization of the clustering algorithm convergence**

**Table 1 The LoRa receiver sensitivity at bandwidth of 125 KHz and the corresponding transmission time for the various spreading factors (SF) and with packet size of 51 bytes**

SF	SLORA at 125 kHz bandwidth	Packet transmission time (ms)
12	-137	1318.912
11	-135	659.456
10	-133	329.728
9	-130	185.344
8	-127	92.672
7	-124	51.456

The results of the minimum and maximum Euclidian distances of sensors within each of the 5 clusters and the mean Euclidian distance of each of the 5 clusters are shown in Table 2. The results in Table 2 show that the overall Average Euclidian Distance (AED) of the sensor nodes in the network is 60.48 m, Cluster 1 has the highest AED with a value of 652.27 m and the sensor node with the highest Maximum Euclidian Distance (SMxED) with a value of 1205.47 m. On the other hand, Cluster 5 has the lowest mean AED with a value of 568.67 m.

The results of the path loss at the minimum and maximum Euclidian distances of sensors within each of the 5 clusters and at the AED of each of the 5 clusters are shown in Table 3. The results in Table 3 show that the overall Average Path loss (APL) of the sensor nodes in the network is 127.405 dBm, Cluster 1 has the highest APL with a value of 128.289 dBm and the sensor node with the highest path loss with a value of 136.3 dBm. On the other hand, Cluster 5 has the lowest APL with a value of 126.289 dBm.

**Table 2 The results of the minimum and maximum Euclidian distances of sensors within each of the 5 clusters and the mean Euclidian distance of each of the 5 clusters**

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Min	Max	Average
No. of nodes	493	321	482	345	359	321	493	400
Percentage of total nodes (%)	24.65	16.05	24.1	17.25	17.95	16.05	24.65	20
Minimum Distance (m)	15.58	113.83	24.57	11.20	58.10	11.20	113.83	44.66
Maximum Distance (m)	1205.47	1078.22	1153.77	1045.90	1161.01	1045.90	1205.47	1128.88
Average Distance (m)	652.27	602.52	643.62	580.67	568.34	568.34	652.27	609.48

**Table 3 The results of the path loss for frequency of 2.4 GHz , 10 dBm fade margin and negligible antenna gains**

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Min	Max	Average
FSPL(dBm) at Maximum Distance (km)	136.3	134.8	135.7	134.4	135.8	134.4	136.3	135.4
FSPL(dBm) at Average Distance (km)	128.289	127.255	128.115	126.774	126.495	126.495	128.289	127.405

Based on the results of the path loss (in Table 3) and transmission time (in Table 2) the transmitter power and transmitter energy were computed. The results of the transmitter power at the minimum and maximum Euclidian distances of sensors within each of the 5 clusters are shown in Table 4. The results in Table 4 show that the overall

Average Transmitter Power (ATP) of the sensor nodes in the network is 21.4 dBm for SF of 7 and 8.4 dBm for SF of 12. Cluster 1 has the sensor node with the highest ATP with a value of 22.3 dBm for SF of 7 and a value of 9.3 dBm for SF of 12.

**Table 4 The results of the required transmitter power for the sensor node with the maximum Euclidian distance in each of the 5 clusters**

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Min	Max	Average
SF	PtX(dBm) at Maxi Dist.	PtX(dBm) at Maxi Dist.	PtX(dBm) at Maxi Dist.	PtX(dBm) at Maxi Dist.	PtX(dBm) at Maxi Dist.	PtX(dBm) at Maxi Dist.	PtX(dBm) at Maxi Dist.	PtX(dBm) at Maxi Dist.
12	9.3	7.8	8.7	7.4	8.8	7.4	9.3	8.4
11	11.3	9.8	10.7	9.4	10.8	9.4	11.3	10.4
10	13.3	11.8	12.7	11.4	12.8	11.4	13.3	12.4
9	16.3	14.8	15.7	14.4	15.8	14.4	16.3	15.4
8	19.3	17.8	18.7	17.4	18.8	17.4	19.3	18.4
7	22.3	20.8	21.7	20.4	21.8	20.4	22.3	21.4

The results of the transmitter energy consumption at the minimum and maximum Euclidian distances of sensors within each of the 5 clusters are shown in Table 5 and Table 6. The results in Table 5 and Table 6 show that Cluster 1 has the sensor node with the highest Transmitter Energy Consumption (TEC) with a value of 8738.5 mw or 39.4 dBm for SF of 12 and a value of 11225.8 mw or 40.5 dBm for SF of 12.

The results of the transmitter energy consumption at the AED of each of the 5 clusters are shown in Table 7. The results in Table 7 show that the overall Average Transmitter Energy Consumption (ATEC) of the sensor nodes in the network is 1447.8 mw or 31.6 dBm for SF of 12.

In all, although the spreading factor, SF of 7 required higher transmitter power than the SF of 12, the energy consumption of SF of 12 is higher than that of the SF of 7 because of the higher transmission time required by the SF of 12.

**Table 5 The results of the transmitter energy consumption in mw for the sensor node with the maximum Euclidian distance in each of the 5 clusters**

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Min	Max	Average
SF	EtX(mw) at Maxi Dist.	EtX(mw) at Maxi Dist.	EtX(mw) at Maxi Dist.	EtX(mw) at Maxi Dist.	EtX(mw) at Maxi Dist.	EtX(mw) at Maxi Dist.	EtX(mw) at Maxi Dist.	EtX(mw) at Maxi Dist.
12	11225.8	7947.2	9777.2	7248.0	10005.0	7248.0	11225.8	9124.6
11	8895.8	6297.8	7747.9	5743.6	7928.4	5743.6	8895.8	7230.8
10	7049.5	4990.6	6139.8	4551.5	6282.8	4551.5	7049.5	5730.0
9	7906.4	5597.3	6886.2	5104.8	7046.6	5104.8	7906.4	6426.6
8	7887.7	5584.0	6869.9	5092.7	7029.9	5092.7	7887.7	6411.3
7	8738.5	6186.4	7610.9	5642.0	7788.2	5642.0	8738.5	7102.9

**Table 6 The results of the transmitter energy consumption in dBm for the sensor node with the maximum Euclidian distance in each of the 5 clusters**

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Min	Max	Average
SF	EtX(dBm) at Maxi Dist.	EtX(dBm)	EtX(dBm)	EtX(dBm)	EtX(dBm)	EtX(dBm)	EtX(dBm)	EtX(dBm)
12	40.5	39.0	39.9	38.6	40.0	38.6	40.5	39.6
11	39.5	38.0	38.9	37.6	39.0	37.6	39.5	38.6
10	38.5	37.0	37.9	36.6	38.0	36.6	38.5	37.6
9	39.0	37.5	38.4	37.1	38.5	37.1	39.0	38.1
8	39.0	37.5	38.4	37.1	38.5	37.1	39.0	38.1
7	39.4	37.9	38.8	37.5	38.9	37.5	39.4	38.5

**Table 7 The results of the transmitter energy consumption in mw and in dBm for the mean Euclidian distance of the sensor nodes in each of the 5 clusters**

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Min	Max	Average
SF	EtX(mw) at Avg. Dist. Per Cluster	EtX(mw) at Avg. Dist. Per Cluste	EtX(mw) at Avg. Dist. Per Cluste	EtX(mw) at Avg. Dist. Per Cluste	EtX(mw) at Avg. Dist. Per Cluste	EtX(mw) at Avg. Dist. Per Cluste	EtX(mw) at Avg. Dist. Per Cluste	EtX(mw) at Avg. Dist. Per Cluste
12	1774.7	1398.7	1705.0	1252.0	1174.1	1174.1	1774.7	1447.8
SF	EtX(dBm) at Avg. Dist. Per Cluster	EtX(dBm) at Avg. Dist. Per Cluster	EtX(dBm) at Avg. Dist. Per Cluster	EtX(dBm) at Avg. Dist. Per Cluster	EtX(dBm) at Avg. Dist. Per Cluster	EtX(dBm) at Avg. Dist. Per Cluster	EtX(dBm) at Avg. Dist. Per Cluster	EtX(dBm) at Avg. Dist. Per Cluster
12	32.5	31.5	32.3	31.0	30.7	30.7	32.5	31.6

#### 4. CONCLUSION

The examined the energy demand for the transmission of packets suing LoRa transceiver based sensor node in a clustered network. The study considered some factors that affect the energy consumption and they included the LoRa spreading factor, the path loss and the transmission time. The impact of path length was also examined by using the Euclidian distance of the clustered sensor modes. The sensor node with the Euclidian maximum distance in each cluster was considered along with the average the Euclidian maximum distance in each cluster. The clustering was done using gap statistics for optimal number of cluster determination and then clustering using K-means algorithm with early centroid determination based-on mean.

#### REFERENCES

1. Agarwal, P., Swami, S., & Malhotra, S. K. (2024). Artificial intelligence adoption in the post COVID-19 new-normal and role of smart technologies in transforming business: a review. *Journal of Science and Technology Policy Management*, 15(3), 506-529.
2. Paiva, S., Ahad, M. A., Tripathi, G., Feroz, N., & Casalino, G. (2021). Enabling technologies for urban smart mobility: Recent trends, opportunities and challenges. *Sensors*, 21(6), 2143.
3. Zhang, J., & He, S. (2020). Smart technologies and urban life: A behavioral and social perspective. *Sustainable Cities and Society*, 63, 102460.
4. Mukhopadhyay, S. C., Tyagi, S. K. S., Suryadevara, N. K., Piuri, V., Scotti, F., & Zeadally, S. (2021). Artificial intelligence-based sensors for next generation IoT applications: A review. *IEEE Sensors Journal*, 21(22), 24920-24932.
5. Osamy, W., Khedr, A. M., Salim, A., Al Ali, A. I., & El-Sawy, A. A. (2022). Coverage, deployment and localization challenges in wireless sensor networks based on artificial intelligence techniques: a review. *IEEE Access*, 10, 30232-30257.
6. Seng, K. P., Ang, L. M., & Ngharamike, E. (2022). Artificial intelligence Internet of Things: A new paradigm of distributed sensor networks. *International Journal of Distributed Sensor Networks*, 18(3), 15501477211062835.
7. Heusse, M., Attia, T., Caillouet, C., Rousseau, F., & Duda, A. (2020, November). Capacity of a LoRaWAN cell. In *Proceedings of the 23rd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems* (pp. 131-140).
8. Sherazi, H. H. R., Grieco, L. A., Imran, M. A., & Boggia, G. (2020). Energy-efficient LoRaWAN for industry 4.0 applications. *IEEE Transactions on Industrial Informatics*, 17(2), 891-902.
9. Ksiazek, K., & Grochla, K. (2021, June). Flexibility Analysis of Adaptive Data Rate Algorithm in LoRa Networks. In *2021*

- International Wireless Communications and Mobile Computing (IWCMC)* (pp. 1393-1398). IEEE.
10. Al-Gumaei, Y. A., Aslam, N., Aljaidi, M., Al-Saman, A., Alsarhan, A., & Ashyap, A. Y. (2022). A novel approach to improve the adaptive-data-rate scheme for iot lorawan. *Electronics*, 11(21), 3521.
  11. Álvarez, G., Fraire, J. A., Hassan, K. A., Cespedes, S., & Pesch, D. (2022). Uplink transmission policies for LoRa-based direct-to-satellite IoT. *IEEE Access*, 10, 72687-72701.
  12. Vogelgesang, K., Fraire, J. A., & Hermanns, H. (2021, December). Uplink transmission probability functions for LoRa-based direct-to-satellite IoT: A case study. In *2021 IEEE Global Communications Conference (GLOBECOM)* (pp. 01-06). IEEE.
  13. Herrería-Alonso, S., Rodríguez-Pérez, M., Rodríguez-Rubio, R. F., & Pérez-Fontán, F. (2023). Improving Uplink Scalability of LoRa-Based Direct-to-Satellite IoT Networks. *IEEE Internet of Things Journal*.
  14. Liang, R., Zhao, L., & Wang, P. (2020). Performance evaluations of LoRa wireless communication in building environments. *Sensors*, 20(14), 3828.
  15. Turčinović, F., Vuković, J., Božo, S., & Šišul, G. (2020, September). Analysis of LoRa parameters in real-world communication. In *2020 International Symposium ELMAR* (pp. 87-90). IEEE.
  16. Kufre M. U. (2022) Analysis Of Lora-Based Iot Sensor Node Required Transmitter Power Variation With Communication Range And Bit Error Performance Configurations. *Journal of Multidisciplinary Engineering Science and Technology (JMEST) Vol. 9 Issue 11, November - 2022*
  17. Roupheal, T. J. (2009). High-level requirements and link budget analysis. *RF and digital signal processing for software-defined radio*, 87-122.
  18. Umargono, E., Suseno, J. E., & Gunawan, S. V. (2020, October). K-means clustering optimization using the elbow method and early centroid determination based on mean and median formula. In *The 2nd international seminar on science and technology (ISSTEC 2019)* (pp. 121-129). Atlantis Press.