An Efficient Model for Cluster-Based Routing in Internet of Things

Ofonime Idongesit Francis, Emmanuel Okoni Bennett, Daniel Matthias

Department of Computer Science, Rivers State University, Port-Harcourt, Nigeria Corresponding author Email: ofonime.francis@ust.edu.ng

DOI: 10.56201/ijcsmt.v10.no3.2024.pg32.44

Abstract

The Internet of Things network is a structure of interconnected gadgets that connects with each other and recognize information about the environment. The integration, network and transmission rules of this network device are highly dependent on the routing protocol used in the architecture. The Internet of Things has become a necessary principle in human life, and routing protocols that aid communication are necessary to realize this vision. This research is an extension to further increase the rate of packet delivery to destination. Earlier research enabled the use of cluster-based routing protocol for energy efficiency forgoing the relevance to traffic pattern which is the sole essence of Internet of Things networking. This research hampers on the packet delivery rate, increase on packets delay, reduction in data collision as well as reduced communication overhead. This research made use of the Anaconda 3 (Python Distribution) as the programming language, MatLab 2023 for the simulation, K-Means algorithm for clustering and designation of master nodes for clusters and the Ad-hoc On-demand Distance Vector (AODV) algorithm for route discovery, packet forwarding and routes maintenance. The idea of this algorithm is to improve the performance of Internet of Things network in terms of packet transmission. The computational time and accuracy level results were calculated using contrasting numbers of sensor nodes and K-values in checking the percent error of each simulation, the accuracy level of clustering was 0.88% and the accuracy of the route discovery was 0.91%. To check the scalability of the network, a comparative analysis was demonstrated with Matlab simulator using various parameters: area size 200m, number of nodes 1000, transmission range 1000m, packet size 5, which gave 87.2% performance, the packet delivery ratio showed an accuracy of 0.98% of high and 0.78% of low, the packet delay ratio showed 0.18% of high and 0.04% of low and the total performance of the system showed an average of 98.4% accuracy in performance.

Keywords: Sensor node, Cluster-Based Routing, Traffic pattern, K-Means Algorithm, AODV Algorithm, Packet delivery, Communication Overhead, Data Collision, Clustering, Route Discovery

INTRODUCTION

The Internet of Things (IoT) is a framework or network of interconnected computing gadgets, mechanical and advanced machines, objects, animals or individuals that are uniquely identified and have the capacity to exchange information over a network without the need for human-to-human or human-computer interaction. The integration, networking and transmission rules of these network gadgets are highly dependent on the routing protocol used in the architecture [1]. The Internet of Things (IoT) has become an indispensable principle of human life. It has become one of the emerging technologies of the modern era, attracting interest from both researchers and industry. With vast technological advancements and the growing popularity of digital assistance in everyday life and work environments that goes along with it, technologies are needed to evolve these application domains to the next level. In order to turn this vision into reality, routing protocols are needed to aid the communication between these things in a decentralized, self-organized and changing infrastructure [2].

Routing can be defined as the selection of the best path possible for computer networks using some predetermined rules. Therefore, in IoT, the best of the routing techniques to use is the cluster-based routing or hierarchical based routing which has to do with the grouping of objects or things, in that the objects or things in that group cluster) are more alike than those in other clusters. A cluster head is chosen from each cluster to maintain the cluster members' information. Clustering is one of the best routing techniques to improve the traffic pattern and energy efficiency of a network.

Cluster-based routing is a routing technique used in Internet of Things networks. In this approach, IoT devices are organized into clusters, where one device acts as the head and communicates with other devices within the cluster. The head device is responsible for the relaying of data between devices within the cluster and to the network. Therefore, it is proper for the devices to communicate and share data and resources amongst themselves without interference, delays and collision. Cluster-based routing enhances the discovery of routes such that once all groups are associated and master nodes chosen based on parameters such as gadget location, function or type of the device, the cluster heads communicate with each other to discover the best routes for the transmission of data [3].

Traffic pattern is one of the challenges or constraints of internet of things, because IoT connects multiple devices, therefore, this research work will focus on the traffic pattern of Internet of Things by using cluster-based routing technique to solve traffic pattern problems such as communication overhead, data collision, unbalanced network loads and network end to end delays.

I. RELATED WORKS

The emergence of numerous technologies, such as machine learning, ubiquitous computing, cheap sensors, and increasingly potent embedded systems, has influenced the development of the Internet of Things. Regarding the areas of energy efficiency, privacy, traffic patterns, and protection in particular, there are many worries regarding the limitations limiting the development of Internet of Things technologies and goods [4]. The interchange of data between people, things, and processes is made possible by the Internet of Things (IoT), which has led to an expanding data

sphere and modern activity models based on a variety of information sources. It is predicted that by 2024, there will be 41.66 billion connected IoT gadgets thanks to cellular IoT and other developing cellular innovations, and by 2025, there will be 5 billion connections through cellular IoT [5]. Routing based on clustering helps in the deduction of traffics in wireless sensor networks by limiting the number of transmission that each of the nodes has to make. This is done by proper traffic pattern of each of the hubs in the network so that the data and resources is properly shared and communicated amongst the nodes. Rather than each hub transmitting information directly to the central station, the information or resources is first forwarded to the master head node, which then sends the aggregated information to the central hub [6]. This reduces the overall size of power needed for each hub to disseminate information and also reduces the overall performance of the network traffic.

According to [7], an Energy Efficient Particle Swarm Optimization (EEPSO) method designed to enable smart master node selection across a variety of IoT gadgets was proposed. An EEPSOC was used to elect a master node after the IoT gadget utilized to discover medical data was clustered. The data was transmitted to cloud servers by the elected master node. As such, the master node was responsible for transmitting IoT data to the cloud through fog devices. Then, to identify the severity of the ailment, an Artificial Neural Network (ANN) classifier was used to diagnose the medical data stored in the cloud. This method worked well for the purpose for which it was intended. Prioritized and energy-efficient (PriNergy) was taken into consideration using the low power and lossy network (RPL) model as a basis, the suggested approach uses the routing protocol for contents-based routing. Before sending data to its destination, each network slot considered network traffic as well as sound and picture data, using a timing pattern. The members of the parent member node in the proposed RPL model can survive a failure as long as the parentless parenthesis is configured and convergence occurs, after which their packets expire owing to time lapse. While the routing functionality of the model was flawless, packet failures and delays were a problem [8].

A routing model was proposed in 2004 with the goal of creating a nearly stateless routing protocol. Data was routed using the hubs' residual energy. The framework was divided into various energy bands by this algorithm, which then creates routing paths. This kind of framework is new and better suited for high-traffic wireless sensor networks since it distributes the load almost evenly among all of the hubs. This technique has the upper hand of removing the constraints in the network, because the hubs will be increasing their energies at a constant measured rate. This mechanism has another benefit of not having to calculate paths every time data is transmitted. Analysis show that the technique achieves fairness and load balancing across the network and significantly increases lifetime, but performs poorly for packet delays [9]. In 2010, [10] proposed a novel routing protocol based on clustering for VANET. They conducted a comparison between their protocol and alternative route discovery protocols, and found that the new standards significantly reduced average routing overhead and average end-to-end delay jitter as the number of vehicles increased. The average end-to-end delay jitter with an increase in the number of vehicles must be small to meet the requirements of real-time traffic applications, which demand a relatively stable data transmission delay time. Although the model was successful, it was limited by a high rate of interference and collisions between data. [11] Proposed the single-hop inter path selection procedure, in which the low energy adaptive clustering hierarchy (LEACH) encourages

a straight exchange from master nodes to the central hub. This procedure is not practical for networks with large regions, though. In the research work, an optimized Orphan-LEACH (O-LEACH) was proposed to help form a novel clustering process that can lead to reduced energy consumption and increased network longevity. The orphan node is equipped with enough energy to try and cover the network. The O-LEACH protocol, which provides the full network's coverage with the fewest number of orphaned nodes and incredibly high connectivity rates, is the main novel contribution of the proposed research.

In [12] proposed a model to provide an enhanced K-Means clustering algorithm, a combination of the largest minimum distance algorithm and the conventional K-Means algorithm was suggested. The traditional K-Means algorithm's problems in identifying the initial focal point can be compensated for by this improved algorithm. The enhanced K-Means algorithm successfully addressed two drawbacks of the conventional algorithm: the first is a stronger reliance on selecting the initial focal point, and the second is an ease of trapping in the local minimum. A Ping Flood violation Pattern Recognition Using a K-Means Algorithm in an Internet of Things (IoT) Network was proposed in [13]. Three distinct scenarios: straight traffic, violation traffic, and combined straight-violation traffic were used in wireless communication experiments. A dataset was created for each scenario. Next, the datasets were divided into two disjoint groups: normal and abnormal. The grouping analysis was generated using the K-Means algorithm. 95931 bytes of data on average were sent in the abnormal group, while 4,068 bytes of data on average were sent in the normal cluster. Using a confusion matrix, the accuracy level of the clustering results was established. With the applied K-Means algorithm, the clustering accuracy was 99.94 percent. The ratios gotten from the confusion matrix were as follows: 98.26 percent for true negatives, 100.01 percent for true positives, 0 points for false negatives, and 1.38 percent for false positives. The model performed as intended, although it had issues with detection accuracy because it was hard to identify the features of the DoS abnormalities.

In [14], a new approach to divide the web configuration as per the node's characteristics was implemented. This procedure was attained via "intelligent transmission". For the purpose of routing the bytes of data, the hierarchical learning-based sectionalized routing (HiLSeR) scheme was used. Hierarchical learning for "topology sectionalization and routing decision making, a multi-dimensional data conduct oriented soft clustering paradigm" was used. The efficacy of the proposed model was evaluated in comparison to alternative schemes during testing. Several constraints were accumulated, including "Energy Unit per Message, Dead node Percentage, Overhead Ratio, Average Latency, and Success Ratio," to show the increased efficacy.

There are numerous ways that clustering can enhance the performance of remote sensor networks. It does this by maintaining the network's traffic amount of work, which lowers the amount of path selection information stored at each sensor's node and the energy dissipated during long-distance transmission [15]. By using master node in clustering, which can further prolong energy, local data can be aggregated and activities can be organized with local participants. As a result, the inactive members may be sleeping or using minimal power. Additionally, clustering as a response to dynamic changes in topology has the benefit of lowering the topography cost of maintenance. In order to be responsive to the changes in dynamic phenomena, structure collaboration needs to be

flexible and configurable. When using a network cluster, topology changes are made solely at the master node level and have no impact on the nearby cluster nodes. As a result, it is possible to greatly reduce the overhead associated with unique topography adaptation [16]. In [17], The CH node in a remote sensor network was to be chosen using the multiobjective Taylor crow optimization (MOTCO) algorithm. It is an amalgam of Taylor series and crow search algorithms. To select the best CH among the different network nodes, the MOTCO algorithm considered the parameters, traffic density, delay, and energy. The network's lifespan was increased by the model, which suggested using a reduction objective function set to select the best node for data transmission. But choosing the best CH node in the network with limited resources is a difficult job. According to [18]), energy effective Artificial Bee Colony (ABC) approach for IoT was suggested. The approach had two phases: firstly the selection of master nodes and secondly the formation of groups. In the first phase the master node is selected using the artificial bee colony method. Then in the second phase the groups are then formatted using the Euclidean distance. The master node collects and exchanges data packets from the group member node to the central hub. Nonetheless, the ABC technique expanded the lifespan of the network; but it faces problems in choosing the master nodes while incrementing the size of rounds in the network.

According to [19], in the context of the Internet of Things, research was conducted on data fusion and anti-collision approaches. Employed to successfully extend the longevity of the sensor network was the enhanced grouping path selection protocol Leach based on distributed data fusion. In order to process the data, fusion and routing technologies were combined. This improved the accuracy and efficiency of data acquisition in sensor networks. A data fusion algorithm based on Gauss membership function was introduced. A novel approach to identifying unknown tags within the reader's function range is the effective binary search anti-collision algorithm known as AEBS, which includes an unknown tag number estimation feature. Empirical research and simulation findings demonstrate that the AEBS algorithm shortens the system's overall recognition time while enhancing the effectiveness and stability of system identification. In [20] they introduced a brand new topographic path selection approach oriented on Energy-Efficient Localization (EEL), which used locality information and the remaining lifespan of SNs to forward data packets to sink nodes. The EEL model effectively adjusted to the dynamic network topological fluctuations and refreshed the locality data of Sensor Nodes (SNs) in an Underwater Wireless Sensor Network (UWSN) at regular intervals. Normalized Advancement (NADV), which established the transmission priority level, was taken into consideration by EEL as it iterated through a group of forwarding nodes of candidates. According to the simulated results, EEL significantly increased the Packet Delivery Ratio (PDR) and decreased the amount of energy used during the routing process, all while effectively locating SNs.

II. SYSTEM DESIGN

A. System Architecture

System Architectural sketch or layout is basically a skeleton layout of a structure depicting components, abstraction levels and other aspects of a software system, the hardware, cloud computing, and network architecture that enable inter-device communication make up the Internet of Things architecture. Administrators can assess, monitor, and preserve system consistency with

the help of the differentiated layers that make up IoT infrastructure. This is depicted in figure 1 below:



Figure 1: Architectural Design of the System

The system presents an approach for an efficient routing technique based on clustering in regards to the traffic pattern of the sensor nodes in the IoT network that makes the routing process more effective to minimize delay of packet delivery, data collision, and communication overhead. This will be carried out in two phases: firstly, the node association and the cluster formation will be done by the K-Means clustering algorithm for the grouping of the nodes and the selection of cluster heads. Secondly, the route discovery, data forwarding and the route maintenance will be done by the AODV routing algorithm.

IoT Sensor Nodes

These are different kinds of gadgets on the Internet of Things (IoT) network that can connect wirelessly to an infrastructure and transmit data. For this research, the IoT devices will be represented with IoT dataset and the dataset used for the implementation of the algorithm is NSL-KDD dataset for IoT sensors.

K-Means Clustering Algorithm

The main idea behind this algorithm is that it helps in partitioning data into K cluster such that each data point belongs to the nearest centroid also helps in reducing the total distances between the data points and the groups that complements them. Once the IoT dataset is preprocessed, this section of the algorithm consists of two phases driven by the K-Means algorithm; they are:

i. Device/Node Association

This is the grouping of the devices into cluster. In the system, k-means clustering algorithm groups the sensor nodes dataset into clusters of similar nodes. In k-means clustering, the data points are shared into various groups sanctioned according to the amount of K. Therefore, if said that K = 5, the datasets are then shared into 5 groups, c1, c2, c3, c4, c5.

ii. Cluster Formation

For the clusters to be assigned a cluster head, the K-Means algorithm executes the following procedures to choose the master node for each cluster.

- Deciding size of the cluster (centroid), K and allocating the set of cluster centroid c1, c2....ck.
- Determining the range among every of the appointed centroid data point.
- Assigning the data point to its closest centroid, which is accomplished by using the following formula to calculate the Euclidean distance:

arg min dist(ci,xi)²

ci€C

Where: dist() is the Euclidean distance.

- After moving the random center point to the real centroid, the actual centroid of the data point for the first group is calculated.
- Continue step iii and iv for the next group, once the group is stable, the algorithm is converged.

AODV Routing Algorithm

Reactive routing protocols, such as AODV, create routes only when necessary and do not maintain a constant routing table for all possible destinations. In this phase of the architecture, the algorithm performs the following action:

i. Route Discovery

In order to exchange data to a receiver node that does not currently have a route to, a source node must first start the route discovery process. An RREQ packet is broadcast to the neighbors by the source node. When a node receives an RREQ packet for a destination it does not know, it can either respond with a Route Reply (RREP) if it knows a route, or forward the RREQ to its neighbors. The RREQ packet includes a sequence number to ensure freshness of route information. A Route Reply (RREP) is generated when the RREQ reaches the destination node or an intermediate node that is aware of a route to the destination. By taking the RREQ's opposite route, the RREP returns to the source node.

Nodes monitor the status of routes they have established. If a link breaks or a node moves out of range, the affected node sends out a Route Error (RERR) message to notify other nodes about the

broken link. The RERR propagates to the affected nodes, causing them to update their routing tables.

ii. Packet Forwarding

Inserting the route into the data packet cache is what the source node does after receiving the RRep packet from the terminus node. The packet is routed so that, once it reaches the destination node, each node indicated in the packet advances it to the following substantiated address.

iii. Route Maintenance

Since nodes stores route information of every route on their route tables, the nodes monitor the status of routes they have established. If a link breaks or a node moves out of range, the affected node sends out a Route Error (RERR) message to notify other nodes about the broken link. The RERR propagates to the affected nodes, causing them to update their routing tables.

B. Network Design Layout

The network design layout of the system depict a 5-layered infrastructure: The application layer which is also called the device or node layer in which all physical device belongs, network layer which is called the transmission layer in which is responsible for the secure transmission of data from the device layer to the information system (base station), base station which is in charge of the administration and provision of interconnection between the clusters. Based on the output from clusters, the base station layer treats, computes and achieves determinations. Figure 2 below depicts the network layout of the system.



Figure 2: Network Layout of the System

III. RESULTS AND DISCUSSION

System Setup

To implement and execute the K-Means algorithm and AODV algorithm, we need a suitable and compatible runtime environment like anaconda Jupiter notebook and matlab simulator installed.

Result Presentation

The following were the results obtained after the clustering and route discovery for data transmission as it relates to the packet delay, overheads, and collision. The system was consequently tested for clustering using the K-means algorithm, the route discovery and data transmission using the AODV algorithm. To evaluate the performance of the system the researcher used the MatLab simulator for the simulation. The accuracy of the system was measured using the percent error. The performance analysis of the system was being compared with the existing system.

Figure 3 showed the total time it took the system to cluster the given datasets and discover a route for packet delivery.

Figure 4 showed the clustering of the sensor nodes using a cluster formation of 200 sensor nodes; the various cluster members with their centroid.

Figure 5 showed the network plotting area for route discovery using average parameters of 100m² for area size, 100 no of sensor nodes, 2 base station, and packet size of 10 for data forwarding.

And finally, figure 6 showed the accuracy margin of the proposed system on various k-sample values using the error percent count.

Also, figure 7 showed the comparative analysis of the system with other existing systems (LEACH, CBRTP, EECB) to determine the performance of the system in terms of packet delivery, packet delays, overhead, and collision.



Figure 3: Total Computational time of the system on the given dataset



Figure 4: K-Means cluster formation with centroid using K-sample values of 200 datasets



Figure 5: Network plotting area for Route Discovery



Figure 6: Accuracy Margin of the system on the given sample values

A. COMPARATIVE ANALYSIS

Figure 7 below shows the comparative analysis that was executed on the system with the existing systems (LEACH, CBRTP, EECB) to determine the performance of the system in terms of packet delivery, packet delays, overhead, and collision.



Figure 7: Packet delivery ratio of data packet sent

S/N	Parameters	Description
1.	Simulation	500m x 500m
	Area	
2.	Simulation	700s
	Time	
3.	Traffic Type	CBR
4.	Data Packet	512 bytes
	Rate	

Table 1: Simulation Parameters

5.	No. of Nodes	50, 100, 200, 500,
		1000, 2000, 3000
6.	Connection	UDP
	Protocol	
7.	No. of Clusters	5, 7, 15, 18, 23
8.	Transmission	50. 100, 150, 200m
	Range	
9.	Grid size	10m
10.	No. of Base	5
	Station	
11.	No. of Packets	2,3,5,7

Table 1 above shows all the parameters used in the Matlab environment for the simulation of the system for route discovery after the nodes were clustered.

IV. CONCLUSION

The system was tested using the Anaconda Jupiter notebook and Matlab simulator and the interface look same in all scenario. The performance of the system in terms of clustering, route discovery and packet delivery is satisfactory. The speed, time taken and accuracy of the system performance is also very satisfactory. The system performed it predetermined function and goals of clustering and route discovery for data and packet forwarding. Although the proposed system was successfully modeled for the purpose of traffic pattern in IoT network, yet there is need for further improvement since the system is lacking in terms of speed when large amount of sensor nodes is clustered for route discovery or large amount of packet is to be forwarded, thus the need for improvement in the areas of larger sensor nodes and packet for delivery, transmission range and system speed (CPU) in order not to become a problem or affect the speed and time of route discovery.

V. **REFERENCES**

[1] Karthick, R., Prahabaran, A. M., and Selvaprasanth, P. (2019). Internet of Things based high security border surveillance startegies. *Asian Journal of Applied Sciences and Technology* (*AJAST*) *Volume 3*, 94-100.

[2] Steenbrink, L. (2014). Routing in the Internet of Things. In L. Steenbrink, *Routing in the Internet of Things* 1. Ausarbeitung: Eingereicht am.

[3] Aravind, K., and Praveen, K. R. (2022). Dingo Optimization Based Cluster Based Routing in Internet of Things. *Advanced Technologies In Sensor Network and Internet of Things*, 20-22.

[4] Sethi, P., and Sarangi, S. R. (2017). Internet of Things: Architectures, Protocols and Applications. *Journal of Electrical and Computer Engineering.*, 1-25.

[5] Yuhong, L., Yuanyuan, H., Xiang, S., and Jukka, R. (2017). Gamma-modulated Wavelet Model for Internet of Things Traffic. *IEEE International Conference (ICC), SAC Symposium Internet of Things Track*, 48-78. Paris.

[6] Qureshi, K. N., Alhudhaif, A., Shah, A. A., Majeed, S., and Jeon, G. (2021). Trust and priority-based drone assisted routing and mobility and service-oriented solution for the internet of vehicles networks. *Journal of Information Security and Applications*, *59*, *102864*.

[7] Bharathi, R. A., Gupta, D., Khanna, A., Elhoseny, M., and Shankar, K. (2020). Energy efficient clustering with disease diagnosis model for IoT based sustainable healthcare systems. *Sustainable Computer Information System*, 100453.

[8] Sachan, F., Souri, T. B., and Aloqaily, M. (2020). PriNergy: a priority-based energyefficient routing method for IoT systems. *The Journal of Supercomputing*, *76*, 8609-8626.

[9] Sasanka, M., Cariappa, M., Rajgopal, K., Arjan, D., and S, S. I. (2004). EBRP: Energy Band based Routing Protocol for Wireless Sensor Networks. IEEE 223-234. LA, USA: IEEE.

[10] Luo, Y., Zhang, W., and Hu, Y. (2010). A New Cluster Based Routing Protocol for VANET. 2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing (pp. 176-180). Wuhan, China: IEEE.

[11] Senthil, G. A., Arun, R., and Kumar, N. (2022). Internet of Things Energy Efficient Cluster-Based Routing Using Hybrid Particle Swarm Optimization for Wireless Sensor Network. *Wireless Personal Communication Article*, 2603-2619.

[12] Youguo, L., and Haiyan, W. (2012). A Clustering Method Based on K-Means Algorithm. Physics Procedia, 1104-1109.

[13] Stiawan, D. M., Mohd, Y. I., Muawya, N. A., Nizar, A., and Rahmat, B. (2021). Ping flood attack pattern recognition using a K-means algorithm in an Internet of Things (IoT) network. *IEEE Access*, 116475-116484.

[14] Banyal, S., Bharadwaj, K., Sharma, D., Khanna, A., and Rodrigues, J. (2021). HiLSeR: Hierarchical learning-based sectionalised routing paradigm for pervasive communication and Resource efficiency in opportunistic IoT network. *Sustain. Comput. Inform. Syst*, 100508.

[15] Abbasi, M. Y. (2007). A survey on clustering algorithms for wireless sensor networks. *Computer Communication*, 2826–2841.

[16] Nikolaos, A. P., Stefanos, A. N., and Imitrios, D. (2013). Energy-Efficient Routing Protocols in Wireless Sensor Networks: *A survey'*, *IEEE Communications surveys and tutorials*, *15*(2), , 551–591.

[17] John, J., and Rodrigues, P. (2019). MOTCO: Multi-objective Taylor Crow optimizarion algorithm for cluster head selection in energy aware wireless sensor network. *Mobile Network and Applications*, 1509-1525.

[18] Yousefi, S., Derakhshan, F., Aghdasi, H. S., and Karimipour, H. (2020). An Energy-Eficient Artificial Bee Colony-Based Clustering in Internet of Things. *Computers and Electrical Engineering*, 86.

[19] Gaurav, K., Himanshu, M., and Akshat, R. S. (2014). An Hybrid Clustering Algorithm for Optimal Clusters in Wireless Sensor Networks. *IEEE Students' Conference on Electrical, Electronics and Computer Science*, , 1–6.

[20] Hao, K., Shen, H., Liu, Y., Wang, B., and Du, X. (2018). Integrating localization and energy-awareness: A novel geographic routing protocol for underwater wireless sensor networks. *Mobile Network Application*, 1427–1435.