

Chapter 20

A Review of WSN Clustering Algorithms for Low Powered IoT Protocols



Wan Nur Hurul Asma Wan Hassan, Aizat Faiz Ramli, Hafiz Basarudin, Izanoordina Ahmad, and Aduwati Sali

Abstract In technological advancement, Internet of things or IoT technology is widely used in everyday life such as environmental monitoring, military, smart lifestyle, and transportation. IoT technology produces a smart device that connects sensors or devices to produce and analyses data. Wireless IoT networks consisting of battery-powered devices require highly energy efficient network access. Large communication links between devices will exhaust energy during the transmission process, causing a shift in network topology and degrading the quality of service QoS. Cluster techniques developed for wireless sensor networks can be applied to IoT to increase energy efficiency, spatial reuse, and large IoT networks' scalability. This paper surveys designed clustering techniques philosophy that can be implemented in IoT.

Keywords Internet of things (IoT) · LoRaWAN · Energy efficiency · Design clustering

W. N. H. A. W. Hassan · A. F. Ramli (✉) · H. Basarudin · I. Ahmad
Electronics Technology Section, Universiti Kuala Lumpur British Malaysian Institute, Gombak, Selangor, Malaysia
e-mail: aizatfaiz@unikl.edu.my

W. N. H. A. W. Hassan
e-mail: whurul.hassan@s.unikl.edu.my

H. Basarudin
e-mail: hafizb@unikl.edu.my

I. Ahmad
e-mail: izanoordina@unikl.edu.my

A. Sali
Computer and Computer System Section, Universiti Putra Malaysia, Serdang, Selangor, Malaysia
e-mail: aduwati@upm.edu.my

20.1 Introduction

The Internet of things (IoT) has gained huge worldwide interest in industry or academia over the past few years. Internet of things is a virtual information network technology including radio frequency detection, sensor, and microcontrollers [1]. Today, home appliance, military, and healthcare equipment has been upgraded to automation systems with minimal treatment. IoT technology achievement develops a simple routine for users to access data from IoT devices anywhere and anytime. IoT popularity has generated business growth opportunities.

It is predicted that there will be over 1.5 billion Internet users in the future, generating a massive demand for Internet-enabled smart objects [1]. According to CISCO research each person in the world will have more than one Internet-connected device [1]. By referring data from the MIMOS Malaysia [9, 19], they indicate an expected IoT market growth from 2020 to 2025. IoT technology has developed various business categories such as smart home, smart city, smart environment, and, etc. [1]. In an IoT platform, communication between any Internet-enabled a computer creates several applications in various domains.

IoT is planned to build and deliver a new generation of service and infrastructure using information and communication technology (ICT) [1]. IoT combines technology and system into one application to enable new lifestyle versatility and performance. Depending on traffic load and power consumption requirements and specifications, IoT networks can be generated solely by wired or wireless communication.

In wireless communication, data is transmitted via the gateway and stored in the cloud before data is sent to the application. IoT is a two-way contact, involving two and more devices [3]. IoT networks consisting of battery-powered devices require highly energy-efficient network access. Communication between devices can exhaust energy during transmission. Node with depleted battery will cause node route change in the IoT network resulting in network topology change and can reduce QoS service network quality. This can affect overhead message resulting in network topology change [3]. Therefore, energy consumption is a concern for wireless IoT network.

In WSN topology, clustering generates a node that is self-organizing multi-hopping network that can function independently during data transmission and can be applied to IoT [4]. Clustering is defined as a process of grouping sum nodes into a community that similarly provides improved data transmission efficiency. Network clustering principle is to boost energy efficiency, reliability, and energy consumption during transmission [4]. Clustering can be implemented in one-to-many, many-to-one, one-to any or one-to-all (broadcast) communication [5]. The cluster consists of three main nodes, cluster head, nodes, and gateway nodes [4]. The cluster head acted as local controller in cluster network. The form of the clusters depends on the route for the nodes to transmit the data. The coverage for the large area can create a time delay, scattering, and data loss during the process.

20.2 Clustering Challenge

Clustering creates a promising approach that can help to minimize many IoT challenges such as energy efficiency, power consumption. Energy efficiency is superior to all IoT applications, but it is highly important for the applications such as smart environmental that do not have access to the power supply. The application lifespan would become uncertain thus it caused a drawback to the network such as nodes failure [4].

Each IoT application need a different type of application requirement. Application such as agriculture monitoring and surveillance are homogenous network architecture that require energy efficiency to distribute through the network.

Figure 20.1 demonstrates the sensor forwarding with and without cluster clustering. Data forwarding without clustering showing distance drawback and the data transmission performance to the base station. Clustering establishes a cluster network to capture node data before sent to base station whether in single-hop clustering or multi-hop clustering. Clustering algorithms build for WSN such as low energy adaption clustering hierarchy (LEACH) and hybrid energy efficient distributed clustering (HEED) can therefore be adapted to IoT [4].

20.3 Methodology

Many forms of clustering were implemented for clustering. Clustering design is influenced by many factors, which include network topology and power consumption. Clustering provides various advantages such as energy efficiency, reduces the communication bandwidth requirements and prevents repeated node data exchange [5]. By stabilizing the cluster topology, the overhead for the topology is reduced [5]. Improving energy efficiency, reduce energy consumption or enhancing transmission reliability still consider as a major challenge to clustering [5]. Cluster head (CH) selection is essential to maintain network lifetime and transmit data to endpoint. The clustering objectives will be elaborated as shown in Fig. 20.2.

One of the main objectives of clustering is to reduce nodes energy consumption during the transmission. LEACH introduces a formation of randomized, adaptive and self-configuring network which minimize the energy of CH during transmission [5]. HEED clustering choose the CH by selecting their residual energy and formation of the equal size cluster to reduce the energy in the network [6]. *K*-means algorithm is known as unsupervised algorithm as the selection of CH is determined by *K* initial to the center point [7]. Each clustering creates an algorithm to create a good energy efficiency in their clustering. Thus, clustering can be used as one option to create a productive network for energy efficiency.

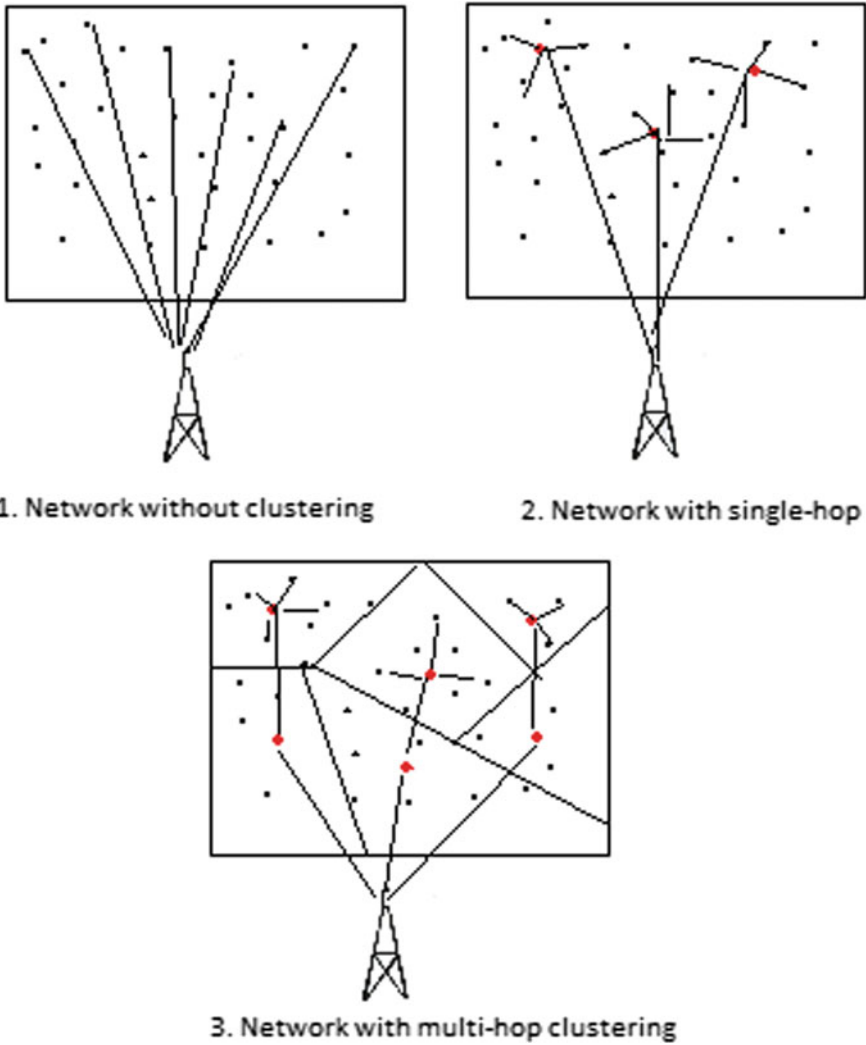


Fig. 20.1 The sensor forwarding data with and without clustering

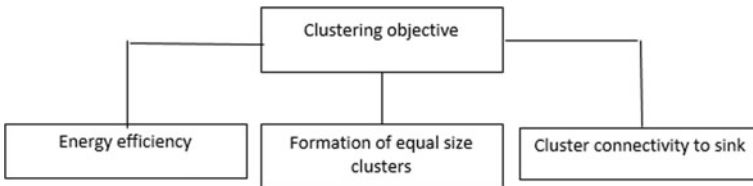


Fig. 20.2 Clustering objective

20.3.1 Low Energy Adaption Clustering Hierarchy (LEACH)

LEACH is a self-organizing adaptive clustering protocol which distributes the energy load equally to the all nodes in the cluster [5]. LEACH is a hierarchical routine which provides a better energy efficiency and scalability in the cluster [8]. There are two phases in LEACH which is the setup phase and the steady-state phase [5]. For this type of protocol, the whole network was divided into cluster and a certain node will be chosen as the cluster head (CH) based on their criteria. The CH will collect, aggregate and compress the data from other nodes in the cluster and transmit to the base station (BS) [8]. The CH acted as lead to the cluster network which consumes more energy compared to other nodes in the cluster. Hence, the energy in the network can be reduced.

The common method used in clustering is the clustering rotation. The rotation is applied in the clustering to balance the energy dissipation within the cluster [8]. The LEACH clustering has two categories which is single-hop and multi-hop communication. The single-hop communication was used in intra-clustering. Intra-clustering and inter-clustering are two type of modes used in LEACH. Intra-clustering, the cluster communicates within the cluster member, but inter-clustering communicates with the neighbor network which outside from their cluster.

The main objective for LEACH is to increase energy efficiency by adapting the rotation CH selection by random number [1]. The CH selection is depending on the value of the threshold $T(n)$ which is given in Eq. (20.1).

$$T(n) = \begin{cases} \frac{1}{1-p(r \bmod \frac{1}{p})} & \forall n \in G \\ 0 & \text{otherwise} \end{cases} \quad (20.1)$$

The value of p is the desired percentage for the nodes to become CH, r denoted the current round of cluster and G is the number not participated in the network. The algorithms calculate the highest $T(n)$ to select the CH in the round [1]. The previous CH will not be allowed to join in the next round to make sure the other nodes to get fair and equally participated to be select as CH. During the steady-state phase, the CH transmits the data by TDMA schedule. During the data transmission, the other nodes go to the sleep mode to consume energy. The energy model for the energy efficiency is written as following [1]

$$E_{Tx}(k, d) = E_{Tx_elec}(k) + E_{Tx_amp}(k, d) \quad (20.2)$$

The channel for the communication is symmetrical and the energy sending will be k bits/packet to a node in distance, d in meter. The E_{elec} is the energy consumed by transmitter or receiver and E_{amp} is the amplifier parameter of transmission correspond to the multi-path fading model. Each packet data bit transmits the information along with overhead. The overhead contains information related to the packet like coding that contain reliable information for transmission. The CH may communicate with the nodes in the cluster depending on the range of transmission. Therefore, to create

a good energy efficiency, the nodes must be within the range of cluster to decide the optimal CH [11].

20.3.2 Hybrid Energy Efficient Distribution Clustering (HEED)

Hybrid energy efficient distributed (HEED) clustering is a clustering process which produces an equal size and the residual energy nodes and intra-clustering communication which plays as significant part in the CH selection [19]. There are three phase of cluster formation in the HEED which are initialization, iteration, and finalization [19]. In the cluster, there will be one node act as CH and in charge to communicate with another CH. The intra-cluster is applied in HEED clustering to increase the energy efficiency and prolong the network lifetime [13]. The energy distribution extends the lifetime of nodes within the cluster, thus stabilizes the neighbor node to operate correctly when nodes are not synchronized.

The selection of CH is determined in the initialization phase. The phase assigns nodes to the probability to become CH. This can be done according to the formula written [19]

$$CH_{\text{prob}} = CH_{\text{prob}} * \frac{E_{\text{residual}}}{E_{\text{max}}} \quad (20.3)$$

The C_{prob} is the initial probability (predefined value), E_{residual} is the residual energy and E_{max} is the maximum energy for sensor nodes [7]. For the interactive phase some nodes will become tentative CH. If the other nodes within the range of tentative CH, it will be tentative CH will become another CH [7]. In the finalization phase, the sensor node which does not have CH will form one cluster.

HEED is an equal size balanced cluster. The HEED clustering used both free space and multi-path channel and consume for the error-free communication link. The calculation for the energy module is written [19]

$$E_{Tx} = (E_{\text{elec}}Xk) + (E_aXkXd^n) \quad (20.4)$$

The E_{elec} is the energy for the transmitter and receiver, E_a is the energy spent by the amplifier and d is the distance between sender and receiver. The transmission energy E_T is the energy to send the data packet.

20.3.3 *K-Means Algorithms*

K-means is a statistical, unsupervised, non-deterministic, iterative strategy for grouping the node into the clusters. It is the simplest unsupervised learning algorithms known for its speed, effortlessness, and usability. *K*-Means utilize the *K* as the initial center point from the consummate dataset [10]. This algorithms objective is to minimum the distance between node to the center point. It calculate the Euclidean distance from each data point, to choose the point which is most suitable, then assign it to the suitable cluster. The center point receives the update and will repeat the calculation until it reaches the minimum to remove the error in the cluster. This algorithms can reduce the impact of isolated point and noise to enhance the efficiency in clustering [8].

K-means is a typical clustering in data mining and can be widely used for large set of data cluster. The algorithms consists of two separate phase [10]. The first phase selects the random *K* center, where the value *K* is fixed. The next phase is to find and take the data nearest the center. The Euclidean distance is considered to determine the distance of object to center point. This process will repeated calculate until it reaches its minimum error in the cluster,

$$E = \sum_{i=1}^k \cdot \sum_{x \in C} |x - x_i|^2 \quad (20.5)$$

20.3.4 *Formation of an Equal Size Cluster*

Formation of an equal size cluster attempts to create a cluster which has a similar number of nodes (cluster members). The large variation of cluster sizes will cause a poor load balancing and thus can affect the performance in certain clusters [12]. The process of creating a more equal size cluster can improve the performance of the network in terms of reducing channel contention thereby reducing latency [12].

20.3.4.1 *Fuzzy Logic Formation*

Fuzzy logic for the formation of uniform size clusters (FUSA) attempts to reduce cluster size variation caused by random placement of nodes in the network. The altruistic decision is used for the nodes to work independently for the selection of CH in the network [14]. The number of neighboring un-cluster nodes in the network is used to determine the selection of CH. The transmission range, *r* between the nodes must be minimized to maintain the network connectivity and energy efficiency. Large cluster size creates low energy efficiency and low transmission coverage which can create data loss during the transaction.

20.3.4.2 Algorithms Cluster Establishment (ACE)

The algorithm for cluster establishment (ACE) plays two parts which are spawning and migrating [15]. The spawn is a node which elects itself to be CH and finding his loyal follower to play the second part. The second part is migrating. The node controls the data migrate to avoid overlap data. During the node iteration, the CH will find his follower to create a path to be synchronized during the data transmission. The iteration reduces the overlap data and avoids repeated path which is used in previous CH leader [14, 15]. The best candidates for CH leaders are the leaders with a loyal follower. This produces a good packing of the repulsion effect between clusters. However, the process to determine the CH leader can create a time delay due to the process of the iteration.

20.3.5 Cluster Connectivity to Sink

Clustering techniques contain two phases which is the cluster head selection and the clustering. The phase remains ongoing until any of the nodes run out of energy. In clustering networks, large wireless sensor are deployed to obtain, monitor the data. Due to this circumstance and limited power limited power consumption of nodes, energy efficiency essentially choose multi-hop communication to create multiple sink [16]. The main modes of communication are multi-hop and single-hop. Single-hop routing is depending on the chain of each other. The chain will be broken when one of the nodes fails to respond. Unlike multi-hop routing which randomly selects the nodes to be the CH in the specific distance to avoid the signal propagation to the BS. The percentage for the chain to be broken is lower than for the single-hop [16].

20.3.5.1 Single-Sink

The cluster was used to cover large area and the single-sink formation is used to transmit the data from nodes to the base station. Single-link minimizes the route of the data and reduces the forward time in the network [11]. Single-hop communication is considerable as straight-forward communication which reduces the time for the data to the base station [17]. However, the main drawback is the lack of redundancy of data.

20.3.5.2 Multi-sink

In the clustering, the network is randomly divided to several clusters. Each cluster is managed by the CH. The sensor node functions as collect data, process data and exchange data with another sensor. A node can be a sink which is generally no energy limitation. Sink collect the information from the sensor nodes [15]. The multi-hop

communication is considerable as an energy efficiency for large scale network [18]. The commonly use for multi-hop topology is the aggregation tree rooted at the sink. The multi-hop topology creates several cluster and CH creates the path for the data to transmit to the sink nodes. However, the uneven cluster size can create unbalanced data distributed which causes the data overloading in the certain CH [18]. Multiple sink are deployed in large scale of network to reduce the redundancy, distribution traffic, and network lifetime. The multi-hop localizes the sink to each cluster to minimize the traffic and redundancy of data [17]. Thus, the combination can create an energy consumption for the transmission in the network.

20.4 Conclusion

In this research, IoT network can be configured based on the cluster criteria such as sink point, CH selection, cluster group to improve data routing efficiency, and network energy consumption. LEACH, HEED, and *K*-Means clustering are an independent network that is efficient in terms of stability based on nodes selection in each neighbor independent of CH selection. Network size defines distance between the nodes and the sink point to minimize data transmission scale. In the technology world, as MOSTI notes, IoT has a positive impact on Malaysia lifestyle and economy. There are a lot of ideas for improvement, as an example for the IoT interface mode during the transmission or network equal size cluster.

Acknowledgements This research is supported by Malaysia Education Minister grant (Application ID-295462-313519).

References

1. Behera TM, Samal UC, Mohapatra SK (2018) Energy-efficient modified LEACH protocol for IoT application. *IET Wirel Sens Syst* 8(5):223–228. <https://doi.org/10.1049/iet-wss.2017.0099>
2. Bhandari S, Sharma SK, Wang X (2017) Cloud-assisted device clustering for lifetime prolongation in wireless IoT networks. In: 2017 IEEE 30th Canadian conference on electrical and computer engineering (CCECE). <https://doi.org/10.1109/ccece.2017.7946815>
3. Bouguera T, Diouris J, Chaillout J, Jaouadi R, Andrieux G (2018) Energy consumption model for sensor nodes based on Lora and LoRaWAN. *J Sens* 18(7):1–23. <https://doi.org/10.3390/s18072104>
4. Cao L, Xu C, Shao W, Zhang G, Zhou H, Sun Q, Guo Y (2010) Distributed power allocation for sink-centric clusters in multiple sink wireless sensor networks. *J Sens* 10(3):2003–2026. <https://doi.org/10.3390/s100302003>
5. Carlos-Mancilla M, López-Mellado E, Siller M (2016) Wireless sensor networks formation: approaches and techniques. *J Sens* 1–18. <https://doi.org/10.1155/2016/2081902>
6. Chan H, Perrig A (2004) ACE: an emergent algorithm for highly uniform cluster formation. *Lect Notes Comput Sci* 154–171. https://doi.org/10.1007/978-3-540-24606-0_11
7. Kumar JS, Zaveri MA (2018) Clustering approaches for pragmatic two-layer IoT architecture. *Wirel Commun Mob Comput* 1–16. <https://doi.org/10.1155/2018/8739203>

8. Li Z, Xin P (2017) Evidence-efficient Multihop clustering routing scheme for large-scale wireless sensor networks. *Wirel Commun Mob Comput* 2017:1–14. <https://doi.org/10.1155/2017/1914956>
9. MIMOS B (2015) National internet of things (IoT) strategic roadmap: a summary (Online). Available: http://www.mimos.my/iot/National_IoT_Strategic_Roadmap_Summary.pdf. Accessed on 10 Sept 2021
10. Na S, Xumin L, Yong G (2010) Research on K -means clustering algorithm: an improved K -means clustering algorithm. In: 2010 third international symposium on intelligent information technology and security informatics. <https://doi.org/10.1109/iitsi.2010.74>
11. Palan NG, Barbadekar BV, Patil S (2017) A retrospective analysis of cluster head selection protocol. In: 2017 international conference on inventive systems and control (ICISC). <https://doi.org/10.1109/icisc.2017.8068724>
12. Perera C, Liu CH, Jayawardena S, Min C (2014) A survey on Internet of things from industrial market perspective. *IEEE Access* 2:1660–1679. <https://doi.org/10.1109/access.2015.2389854>
13. Rajan AA, Swaminathan A, Brundha, Pajila B (2019) A comparative analysis of Leach, teen, Sep and DEEC in hierarchical clustering algorithm for WSN sensors. *Lect Notes Comput Sci* 395–403. https://doi.org/10.1007/978-3-030-28364-3_39
14. Ramli A, Basarudin H, Abu MA, Yaakop M, Sulaiman MI (2017) FUSA: fuzzy logic based clustering protocol for formation of uniform size clusters. In: 2017 international conference on engineering technology and technopreneurship (ICE2T). <https://doi.org/10.1109/ice2t.2017.8215997>
15. Sholla S, Kaur S, Begh GR, Mir RN, Chishti MA (2017) Clustering internet of things: a review. *JST* 3(2):21. <https://doi.org/10.31130/jst.2017.61>
16. Singh SK, Kumar P, Singh JP (2017) A survey on successors of LEACH protocol. *IEEE Access* 5:4298–4328. <https://doi.org/10.1109/access.2017.2666082>
17. Ullah Z, Mostarda L, Gagliardi R, Cacciagrano D, Corradini F (2016) A comparison of HEED based clustering algorithms—introducing ER-HEED. In: 2016 IEEE 30th international conference on advanced information networking and applications (AINA). <https://doi.org/10.1109/aina.2016.87>.
18. Xu L, Collier R, O’Hare GM (2017) A survey of clustering techniques in WSNs and consideration of the challenges of applying such to 5G IoT scenarios. *IEEE Internet Things J* 4(5):1229–1249. <https://doi.org/10.1109/jiot.2017.2726014>
19. Younis O, Fahmy S (2004) HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *IEEE Tran Mob Comput* 3(4):366–379. <https://doi.org/10.1109/tmc.2004.41>