A Distributed Energy-Efficient Unequal Clustering based Kruskal Heuristic for IoT Networks

Mohamed Sofiane Batta\textsuperscript{1,2}, Zibouda Aliouat\textsuperscript{2}, Hakim Mabed\textsuperscript{1}, and Malha Merah\textsuperscript{2}

\textsuperscript{1} FEMTO-ST Institute/DISC, University of Bourgogne Franche-Comte, Montbéliard, France
\textsuperscript{2} LRSD Laboratory, Computer Science Dept., Ferhat Abbas University Setif 1, Setif, Algeria

\textbf{Abstract}

Energy efficiency is a major concern and a critical issue for energy constrained wireless networks. In this context, clustering is commonly used for topology management and maximizing the network lifetime. Clustering approaches typically use a multi-hopping mechanism where Cluster Heads (CHs) near the Base Station (BS) consume higher energy since they relay data of farther CHs. Therefore, nodes close to the BS are strangled with an overloaded routing task and tend to die earlier than their intended lifetime, which affects the network performance. This situation is known as the hot spot problem that induces unbalanced energy consumption among CHs. The concern in this work is to address the intra-clustering structure in large scale environments to tolerate the network scaling and reasonably balance the energy consumption among CHs. In this regard, we propose a new Unequal Clustering algorithm based on Kruskal heuristic (UCKA) to optimize the network lifetime. UCKA applies the Kruskal heuristic in a distributed fashion to perform a minimum spanning tree within large cluster which strengthen the intra-cluster routing structure and reduce the energy devoted to wireless communications. To the best of our knowledge, this is the first solution that combines the Kruskal heuristic and the unequal clustering to extend the devices durability and alleviate the hot spot problem. Simulation results indicate that UCKA can effectively reduce the energy consumption and lengthen the network lifetime.

\textbf{Keywords}

IoT, WSN, Energy-aware protocols, Unequal Clustering, Hot spot energy problem, Kruskal Heuristic.

\section{Introduction}

The efficient use of the Internet of Things (IoT) and Wireless Sensor Networks (WSNs) enables a practical solution for various applications and impacts several daily life aspects. These networks have become a point of interest for many researchers due to their implementation in various real-world scenarios from environmental monitoring and traffic surveillance to smart city and healthcare systems. These networks are made of a large number of intelligent devices, known as connected devices, distributed over a large geographical area and cooperatively interact to perform a specific task.
The communication with the external world is made using a gateway called Base Station (BS). Each device represents a limited on-board processing, limited memory and low radio capacity. Besides, these devices are usually powered by non-rechargeable batteries. Therefore, due to the extensive tasks handled by network devices, energy awareness is a major concern and a critical design issue for wireless networks. To strengthen the energy efficiency and increase the network scalability, the IoT network is commonly organized into clusters. Several clustering protocols are proposed in the literature [5,6,16]. Network devices are grouped into clusters, with a representative Cluster Head (CH) in each cluster responsible for collecting and aggregating data of Cluster Members (CMs) as shown in Figure 1.

![Figure 1. Network clustering topology.](image)

The network clustering is a popular energy efficient technique to improve the energy consumption [9]. In this technique, multi-hop communications between the data sources (CHs) and the gateway (BS) are used to preserve the energy. However, with multi-hop routing, CHs cooperate with each other to forward the data to the base station (BS). Consequently, CHs closer to the BS are burdened with heavy relay traffic of other CHs and tend to die earlier [18]. This scheme reduces the network coverage (sensing area for example) and may cause a network partitioning, especially in large scale networks. Proposed clustering approaches usually generate clusters of even size [21]. Therefore, nodes closer to the BS waste much more energy during the inter-cluster routing because they have a higher load of relaying remote CHs traffic. Thereby, they deplete their energy quickly. This issue is known as the hot spot problem [8]. One solution to mitigate this problem is the use of unequal clustering method, where the network is partitioned into uneven clusters as in Energy-efficient Multi-hop routing with Unequal Clustering (EMUC) [11] and Multi-hop Low Energy Adaptive Clustering Hierarchy (MH-LEACH) [1].
Several unequal clustering proposals have been proposed to extend the network lifetime [19, 22]. However, the majority of these approaches [2, 8] consider small networks and focus on controlling the size of each cluster to balance and reduce energy dissipation, whereas, the intra-clustering structure of the generated clusters is not considered. Moreover, in most cases the connections between the CMs and their CH are assumed to be direct, which may obstruct the communication performances when the network scales (due to the collisions and interference problems [4]). In this work we propose an unequal clustering approach to alleviate the hot spot problem in large scale networks. In the proposed scheme, clusters size varies proportionally to the distance toward the BS. Clustering topology is generated using the Kruskal spanning tree heuristic [14] to reduce the intra-cluster communication cost and expend the cluster lifespan. The objective of integrates the Kruskal heuristic in the unequal clustering is to alleviate the hot spot problem and optimize the network lifetime. Both residual energy and the nodes density are considered in the clustering process to strengthen the energy efficiency.

The proposed approach aims to ensure the load balancing of the energy consumption among CHs by varying clusters size. Clusters, close to the BS, represent small and equal size, where cluster members are directly linked to the corresponding CH. In this case, the CH communicates directly with the BS. Whereas, remote clusters are larger and cluster members are organized into a spanning tree where the CH is the root. The spanning tree is generated using the Kruskal heuristic and defines the intra-cluster communication links between the cluster members. Distant CHs relay their cluster data to the BS through multi-hop transmissions between CHs. With this scheme, smaller clusters are assigned for CHs nearby the BS since they act as a router node for other far CHs. On the other hand, distant nodes spent their energy for multi-hop intra-cluster transmissions.

The rest of this paper is organized as follows: Section 2 describes some related works. Section 3 presents the contribution. Simulation settings and results are discussed in Section 4. We conclude our work in Section 5.

2. RELATED WORKS

The lifespan of sensor nodes is restricted by the lifetime of their batteries. To preserve the energy and increase the lifespan of the wireless devices, many clustering algorithms have been proposed [2,7,8,21]. In this section, some of the relevant works in this field are reviewed.

Low Energy Adaptive Clustering Hierarchy (LEACH) algorithm [13] is one of the most prevalent clustering techniques. To reduce the energy consumption of nodes and enhances their lifetime, each node computes a probability value of subsequently becoming a CH during the next round. The main disadvantage of LEACH is that some CHs may be close to each other. Besides, LEACH uses one-hop communication architecture, i.e. CHs are directly connected to the BS. Several LEACH versions [1,15] attempt to overcome LEACH drawbacks and improve this scheme. Nie et al. [8] proposed a Lifetime-Aware Clustering approach based on a Directed Acyclic Graph (DAG) algorithm. The authors used a linear programming approach to construct a connected routing tree within the network. Xia et al. [23] presented a distributed energy-efficient approach based on Unequal Clustering and Connected Graph theory (UCCGRA). A voting mechanism is used to select the final set of CHs. The clustering takes into count the residual energy of nodes and the distance between the neighbours to elect the cluster heads. A Connected Graph Theory (CGT) is used for inter-cluster communication in order to reduce energy consumption. CHs form a connected routing tree with the BS as root. Therefore, CHs maintain many paths to reach the BS which improves the reliability of transition.
An Unequal Multi-hop Balanced Immune Clustering protocol for wireless networks (UMBIC) is presented in [20]. It uses an Unequal Clustering Mechanism (UCM) and a Multi-Objective Immune Algorithm (MOIA) to avoid the hot spot problem. UMBIC aims to strengthen the network coverage, by producing an optimized routing and ensures low communication cost among nodes. The CH replacement occurs when the residual energy of the CH becomes lower than a particular threshold value. Authors in [12] presented a new clustering approach called Energy Degree Distance Unequal Clustering Algorithm (EDDUCA) to approximate the equalization of energy consumption in a wireless network and eliminate the hot spot problem. The network is divided into unequal clusters by using the Sierpinski triangle method [12]. The clustering is based on the residual energy, node degree, and distance toward the BS. A weight is associated with each node, which is calculated based on the above three criteria. In each cluster, the node with the maximal weight is selected as CH. The objective of EDDUCA is to extend the lifetime of CHs closer to the BS and maintain the clusters’ connectivity.

Baranidharan et al. [3] introduced a new Distributed Unequal Clustering algorithm using Fuzzy logic (DUCF). A Fuzzy Inference System (FIS) is adapted for the CHs election and the energy load balancing. The FIS system considers the residual energy, node degree, and distance to the BS as input and assigns a maximum limit of CMs to each cluster. DUCF uses a Centroid method for defuzzification. The FIS specifies the probability of each node to become CH and determines the size of each cluster according to the input parameters. DUCF assigns a maximum limit of CMs for a CH based on its residual energy and number of neighbours in order to bypass the hot spot problem. To extend clusters lifetime, authors in [10] proposed an Energy Efficient intra-clustering technique (EE3C) where multiple high energy nodes act as CHs within each cluster instead of a single cluster head. The BS collects nodes’ information to determine their location and lifetime. Then it divides the network into rectangular sectors (clusters) to efficiently distribute the energy in the monitored area. The CH election is centralized at the BS. Only one cluster head acts as master CH for a given cluster to send the collected information to the BS. The master CH changes periodically among CHs after a particular number of rounds. EE3C improves the intra-clustering efficiency, however, the clustering process is fully centralized. The authors also presented a k-hop Energy Constrained intra-clustering technique based on the Dominating Sets theory called K-ECDS. The proposed algorithm takes into account the energy limitation. Besides, K-ECDS models the problem of choosing cluster heads using the quality of communication channels and neighbours cardinality.

Unequal cluster-based protocols can mitigate the hot spot problem. However, existing schemes focus on managing the clusters size and do not consider the intra-clustering structure of the generated clusters, which may restrict the scalability in case of a large scale network. Thereby, in our proposed approach, the intra-clustering topology is considered and the proposed scheme is evaluated under a large network with different nodes density to cover several use-case scenarios.

3. THE PROPOSED SCHEME

The wireless network is mapped into a graph $G(V, E)$, where $V$ represents the set of nodes (sensor devices) and the edges $E$ constitutes the communication links. The set of neighbouring nodes of a node $i$ is described by $N(i) = \{ j \in V \mid (i, j) \in E \}$. The distance (number of hops) between two nodes $i$ and $j$ is represented by $D_{i,j}$. The transmitting range is denoted by $T_x$. The proposed scheme uses a weight-based technique that selects the node with the maximum weight in its neighbourhood as a CH. Indeed, in addition to designing an adequate cluster, the CHs selection criteria is crucial for balancing the energy usage among the entire network in a way that all devices operations finish at approximately the same time, i.e. this criteria ensures that the network lifespan is optimized [18].
The nodes weight is calculated based on the following two parameters.

1. **Remaining energy** \((RE_i)\): As the CH performs more tasks than ordinary nodes, it requires more energy. Therefore, the residual energy \(RE_i\) of each node \(i\) is considered during the CHs election. To estimate the battery charge diminution, we used the energy consumption model proposed in [7].

2. **Node density** \((\delta_i)\): Due to the aggregation and interference problems, the energy consumed by a node increases according to its degree. Thus, this parameter is taken into consideration. The neighbourhood density is computed using the received signal strength [24] of adjacent nodes.

Based on the previous parameters, the weight of a node \(i\) is computed as in Equation 1.

\[
W_i = \{\alpha \times RE_i + \beta \times \delta_i\} \land \alpha + \beta = 1
\]  

\(\delta_i\) represents the degree of the node \(i\). \(\alpha\) and \(\beta\) represent the weighting coefficient of the two criteria: residual energy and node degree respectively. \(\alpha\) and \(\beta\) are chosen regarding the target application and the surrounding environment.

A particular weight coefficient may be adjusted relatively to the others to acquire an optimal result for a particular network configuration. For example, in a low density environment, the residual energy should be favoured. Whereas, in case of a dense network, the connectivity should be considered. The proposed approach is designed to operate under a typical network with different configurations to cover various use-case scenarios. Hence, in this experiment, the weight coefficients are considered equal \((\alpha = \beta = 0.5)\).

In the proposed scheme, the network is divided into three types of unequal clusters: small, medium, and large clusters. The size of a cluster depends on the distance between the CH and the BS. Clusters close to the BS are assigned a small size. Hence, CHs of these clusters regroup a reduced number of members, which reserves their energy for routing remote CHs data. In small clusters, intra-cluster communication is done by direct transmission between the cluster member node (CM) and its CH. Medium and large clusters are two and three times larger than small clusters. As the size increases, clusters tend to regroup more members, and direct transmission between the CMs and their CHs may not be feasible. That is why medium and large clusters employ the two-hop and three-hop intra cluster communications to route CMs data to the corresponding CH. Figure 2 shows an overview of the UCKA clustering topology.

Network nodes determine their Cluster Range \(CR_i \in \{small, medium, large\}\) using Equation 2.

\[
CR_i = \begin{cases} 
\text{Small} & \Omega_i \leq 1/3 \\
\text{Medium} & 1/3 < \Omega_i \leq 2/3 \\
\text{Large} & \text{otherwise}
\end{cases}
\]

\(\Omega_i = \frac{D_{MAX} - D_{i,BS}}{D_{MAX} - D_{MIN}}\)

\(D_{MIN}\) and \(D_{MAX}\) are estimated by the BS, they represent the distance between the nearest and the furthest node with respect to the BS. In order to improve the energy efficiency of routing within clusters, medium and large clusters use the Kruskal spanning tree to design the intra-cluster topology.
The Kruskal heuristic computes the minimum spanning tree within each cluster. For this purpose, each CM constitutes initially a singleton tree. Then at each stage, separated trees come together in pairs. The connection between the isolated trees is made using the shortest path that relays these trees (without forming a cycle). Successively, the process is repeated until connecting all the trees in the cluster. The heuristic result is a single minimum spanning tree (MST) obtained using successive connections between isolated trees. The MST relays each node to its CH using the shortest routing path. The clustering procedure is described in algorithm 1.

**Algorithm 1 UCKA clustering process**

**Begin**

**STEP 1:** Compute the node weight \( W_i \) using equation 1.

**STEP 2:** Determine the cluster range \( CR_i \) using equation 2.

**STEP 3:** If the current node has the greatest weight among the neighbourhood then

\[
W_i = \text{Max}(W_j \mid \forall j \in N(i))
\]

Broadcast a \( CH\_Announcement \) message

Else

Upon receiving \( CH\_announcement \) Do

Send a join request \( CM\_join \) to the CH with the highest weight.

Upon receiving a reject message \( CM\_reject \) Do

Repeat STEP 3

**STEP 4:** Upon receiving \( CM\_join \) from node j Do

Update list of cluster members \( CM\_list \)

\( CM\_list = CM\_list \cup j \)

**End**
The procedure of the MST formation is centralized at the CH and is illustrated in algorithm 2. This later may require an additional time to form the clusters. However, the generated structure optimizes the energy consumption and allows the tolerance of the dynamic network topology. Figure 3 shows an execution example of the MST formation.

**Algorithm 2 Intra-cluster MST formation**

**Begin**

\[ T = \emptyset \] //contains the set of sub-trees in the cluster.

\[ S = \emptyset \] //contains the cluster edges.

**STEP 1**: each node in the cluster constitutes a singleton tree.

\[ \forall i \in CM\_list \text{ Do} \]

\[ \text{Tree}_i = \text{Create\_tree}(i) \]

\[ \text{Add Tree}_i \text{ to } T: \]

\[ T = T \cup \text{Tree}_i \]

**STEP 2**: Create a set \( S \) containing all the edges in the cluster \( \forall (i, j) \in E \land (i, j \subset CM\_list): S = S \cup (i, j) \)

**STEP 3**: Connect all the sub-trees in \( T \)

\[ \text{While } S \neq \emptyset \text{ Do} \]

\[ \text{Tree}_i = \text{Find\_Tree}(i, T) \]

//Find\_Tree(i, T): return the tree in T to which node i belong.

\[ \text{Tree}_j = \text{Find\_Tree}(j, T) \]

If Tree_i ≠ Tree_j then

If (i,j) connects Tree_i, Tree_j without forming a loop then

\[ \text{Tree}_z = \text{Merge}(\text{Tree}_i, \text{Tree}_j) \]

\[ \text{Remove} \text{ Tree}_i, \text{ Tree}_j \text{ from } T \]

\[ \text{Add} \text{ Tree}_z \text{ to } T \]

endIf

endIf

Remove (i,j) From S

Endwhile

**End**

4. **PERFORMANCE ANALYSIS**

The simulation experiment is performed using Java Universal Network/Graph (Jung) [17] in Eclipse platform. Jung is a Java-based library that allows modelling and displaying a wireless network as a graph. The performance of UCKA is evaluated against two similar (in terms of objectives) clustering protocols, namely, EMUC [11] and MH-LEACH [1]. Sensor nodes are randomly deployed over a large area (1000 × 1000 m²). Communication links are symmetric and the transmitting range of all nodes \( T_x = 90 \) m. The initial energy of devices is set to 1 joule. For the performance evaluation of the proposed protocol, three metrics are used, notably, the number of clusters generated, the average energy consumed, and the network lifetime. Simulation results are discussed in the following subsections.
Figure 3. MST execution example.

4.1. Average number of clusters

Figure 4 compares our UCKA algorithm with EMUC and MH-LEACH methods in terms of generated clusters number. We observe that the topology generated by UCKA presents fewer clusters compared to MH-LEACH and EMUC. This is due to the fact that UCKA considers the cardinality of nodes during the clustering process and chooses the cluster size according to the distance toward the BS. Whereas, MH-LEACH does not consider nodes density. EMUC takes into account the cardinality of nodes during the CH election, but the size of the cluster only depends on the transmission range of the CH. Therefore, when the network density increases, the number of clusters generated by EMUC tends to increase. Globally, the proposed scheme shows an average improvement of 43% and 34% compared to EMUC and MH-LEACH, respectively.
4.2. Average consumed energy

Figure 5 shows the average energy consumed according to the network density. It shows that the proposed scheme has better performances compared to MH-LEACH and EMUC. The energy efficiency expresses the benefit of the clustering process. As CHs consume more energy compared to other nodes, reducing the set of selected CHs and the use of unequal clustering have balanced and reduced the energy consumption. Indeed, the proposed approach reduces the energy consumption by an average of 24.5% and 8.7% compared to EMUC and MH-LEACH respectively.
4.2. Network lifetime

In this experiment, the network lifetime is defined as the time duration until the last node in the network died (measured in rounds). It can be observed in figure 6 that nodes using the proposed approach stayed alive for a longer period compared to other approaches. This is due to the usage of the unequal clustering model which enables the load balance of tasks among nodes and improves the energy solicitation. In addition, the usage of Kruskal spanning-tree method has improved the energy consumption inside the clusters. Consequently, UCKA has improved the lifespan of nodes and extended the network durability by an average of 41.8% and 12.8% compared to MH-LEACH and EMUC respectively.

5. CONCLUSION AND FUTURE WORK

The hot spot energy problem represents one of the main limits of wireless networks. Indeed, devices close to the base station experience a higher data relay load, which exhaust their energy and leads to an earlier death comparing to other devices. In this paper, we presented an energy efficient unequal clustering approach based on the Kruskal heuristic to minimize the energy dissipation and mitigate the hot spot problem in large scale IoT networks. The proposed scheme considers the residual energy and the network density in the clustering process. Moreover, it reasonably balances the energy consumption over network clusters which in turn optimize the network lifetime. Simulation results show that the proposed scheme is effective in terms of energy and network durability. It reduces the energy consumption by an average of 16% and extends the network lifetime by an average of 27% compared to similar approaches in the literature. In future works, we aim to integrate a deep learning technique for the CHs election to further extend the network durability in the long-term. We aim to investigate more factors, such as the state of health of devices batteries to further improve the performance of UCKA.
REFERENCES


Mohamed Sofiane BATTA received the Master 2 degree in networks and distributed systems from the University of Ferhat Abbas Setif-1, Algeria, in 2018. He is currently pursuing the Ph.D. degree with the LRSD laboratory, Setif, Algeria and the Femto-ST/DISC laboratory, Montbeliard, France. He is working on the field of networks and distributed systems. His research interests include wireless sensor networks, wireless communication management in the IoT networks, dynamic Ad-Hoc networks.

Zibouda ALIOUAT obtained her engineer diploma in 1984 and M.Sc. in 1993 from Constantine University. She received her Ph.D. from Setif University of Algeria. She was an assistant at Constantine University from 1985 to 1994. She is currently a Professor in Computer Engineering Departmentat Setif University of Algeria. Her research interests are in the areas of wireless mobile networks modelling and simulation, wireless sensor networks and fault tolerance of embedded systems.

Hakim MABED is an associate professor at the University of Franche-Comte (UFC), France. He is part of the FEMTO-ST institute (UMR CNRS 6174) and the complex networks team where he does his research. He obtained the Ph.D. degree from the University of Angers, France in 2003; he received the M.S. degree from the USTHB, Algeria in 2000. His research interests are in distributed intelligent MEMS, optimization, distributed algorithms, self-reconfiguration, and mobility.

Malha MERAH received the bachelor’s degree (2018), and the master’s degree (2020) in Computer Science from Ferhat Abbas University of Setif-1, Algeria. She is currently a Ph.D. student at Ferhat Abbas University of Setif-1, Algeria. Her research interests are concerned with the Internet of things (IoT), wireless communications, routing and energy efficiency using machine learning techniques.