Scalability Analysis of the UDCOPA Protocol in Large and Massive IoT Environments

Foudil Mir

Medical Computing Laboratory (LIMED), Faculty of Exact Sciences, University of Bejaia, 06000 Bejaia, Algeria, foudil.mir@univ-bejaia.dz

Farid Meziane

Data Science Research Centre, University of Derby, UK, f.meziane@derby.ac.uk

Clustering has a very positive impact on any optimization problem on the Internet of Things (IoT) or Wireless Sensor Networks (WSN). Energy efficiency based on clustering has proved its efficiency in this area of research, increasing the lifetime of the network and the high availability of services provided by applications based on this type of networks. Unequal clustering represents an advance on equal clustering in terms of flexibility, as it does not impose a predefined radius for clusters formation by elected CHs. Instead, CHs can dynamically adjust the size of their clusters or the radius of their condidature or election according to various factors and criteria, such as energy constraint. As a result, this type of clustering optimizes energy consumption, balances the load between CHs and improves scalability. Unequal-DCOPA (UDCOPA) is an unequal clustering protocol, which enhances the DCOPA protocol (A Distributed Clustering Based on Objects Performances Aggregation for Hierarchical Communications in IoT Applications), that allows CHs to optimize their energy and send a message announcing their solicitation over an Adaptive Radius of Clustering (ARC) that is adjusted according to its local parameters. In this paper, we explore the geographical and quantitative scalability of this protocol as well as the load balancing of clusters and CHs. The results show that UDCOPA is a scalable protocol that maintains its energy and lifetime properties even in geographically very large areas and in massive environments.

CCS CONCEPTS • Networks•Networks~Network protocols~Application layer protocols

Additional Keywords and Phrases: IoT, UDCOPA protocol, unequal clustering, multicriteria analysis, scalability, load balancing, energy efficiency.

ACM Reference Format:

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

1 Introduction

IoT and WSN are completely changing our world by connecting billions of intelligent objects and generating an exponential flow of data. WSNs play a crucial role in collecting environmental data and providing strategic information in various fields such as environmental monitoring, infrastructure management, intelligent agriculture, intelligent transport, connected health and many others. Lately, electric vehicle technology [1][2] offered an interesting application domain on WSNs allowing their integration and the creation of intelligent systems. This explosive growth poses major challenges in terms of network management and optimization, particularly in terms of energy efficiency and scalability. Clustering protocols offer a powerful approach to meeting these challenges. By grouping network nodes into clusters and designating CHs to coordinate communication, these protocols can improve resource utilization and extend network lifetime. What sets UDCOPA distinct from other clustering protocols is its energy management approach based on an adaptive communication radius. This radius is dynamically adjusted according to the local conditions of the CH nodes, allowing CHs to reduce their energy consumption and improve clustering performance.

The aim of this research is to evaluate the geographical and quantitative scalability of UDCOPA. Our objective is to determine whether UDCOPA remains a high-performance protocol when deployed in large-scale environments. To this end, we conduct an in-depth study in three stages. (i) Evaluation of geographical scalability; We simulate the behavior of UDCOPA on networks of different geographical areas. We analyze the impact of extended networks on energy efficiency and lifetime. (ii) Quantitative scalability evaluation; We study the behavior of UDCOPA in the presence of an increasing number of nodes in a network, Hence, evaluating the performance of UDCOPA in massive networks. We analyze the impact of network density on energy efficiency and network lifetime. (iii) Exploring load balancing of the number of nodes distributed on each cluster and load balancing in the geographical extent of clusters and CHs. The results of this research will provide a better understanding of the potential and limitations of UDCOPA in large-scale contexts. This research work addresses a crucial problem in the field of IoT and WSN, which is scalability. By evaluating the scalability of UDCOPA, we hope to make a significant contribution to understanding how the protocol operates in massive and extended environments. These analyses and studies will enable us to open up new solutions based on this protocol.

1.1 Motivations

In this section, we present the main motivations that led us to carry out this work. The exponential growth of IoT and WSN requires efficient resource management to ensure stable and sustainable connectivity. Faced with the challenges of energy efficiency and scalability, clustering protocols offer a promising solution by grouping nodes into clusters to optimize resource use and extend network lifetime. UDCOPA stands out for its approach based on an adaptive communication radius, offering a potentially more efficient solution than traditional protocols. Evaluating its geographical and quantitative scalability in large environments will help to measure its effectiveness and encourage its adoption in real applications. This research aims to address the crucial issue of scalability in these domains, making a significant contribution to understanding the performance of UDCOPA in complex environments.

1.2 Research Questions

In our contribution, the research questions that guided our approach are formulated as follows.

- How does the UDCOPA protocol support geographical scalability and the variety of geographical surfaces in IoT and WSN networks?
- What impact do large networks have on the energy efficiency and lifetime of UDCOPA?
- How does UDCOPA maintain its performance, mortality rate, load balancing, clusters and CH distribution, in massive IoT and WSN environments, and what strategies does it use to maintain optimal efficiency despite scaling in terms of number of nodes or monitoring area?

2 RELATED WORKS

This section is devoted to a review of the few related works focusing on the concept of unequal size clustering in WSN. We discuss some algorithms and protocols proposed in the literature to dynamically adapt and adjust the clustering radius according to some specific and relevant criteria.

UDCOPA, proposed by Mir et all in [3], improved the equal clustering protocol DCOPA proposed by the same authors in [4][5], allows to calculate a clustering radius with the principle of the multicriteria aggregation of the two criteria of residual energy and distance to the Base Station (BS). With this principle, different radius of each elected CH will be defined according to their own proprieties.

The HUCL [6] (Hybrid Unequal Clustering with Layering) protocol aims to optimize network lifetime by overcoming the drawbacks of existing clustering protocols, which are either static or dynamic. By dividing the network into layers and offering clusters of different sizes, with a minimized clustering frequency, the HUCL protocol combines the advantages of both approaches. The installation phase includes the discovery of neighboring nodes, the competition of CHs, the formation and optimization of clusters, and the construction of the transmission path to the BS. The transmission phase uses TDMA scheduling to enable the CHs to transmit the aggregated data efficiently. Although the HUCL protocol addresses some of the challenges of both static and dynamic approaches, it does not take into account the number of neighbors in the calculation of the radius of competition, which is a limitation to consider.

The DSBCA (Balanced Clustering Algorithm with Distributed Self-Organization) protocol, developed by Liao et al. [7], introduces a balanced clustering mechanism based on distance from the BS and node density. Cluster size varies according to these parameters, with larger radius for nodes further from the BS and lower density, and smaller radius for nodes closer to the BS with higher density. The protocol comprises three phases: CH selection, cluster construction and a re-election cycle. In the CH selection phase, random nodes are chosen as candidate CHs, elected on the basis of density and energy. The cluster construction phase limits the size of the clusters to avoid the formation of large clusters. The cycle phase restarts the clustering process after each interval. DSBCA load-balances the clusters by forming different clustering layers, thus localizing the selection of CHs in the same cluster to reduce communication costs.

Energy-Aware Distributed Unequal Clustering (EADUC), presented by Yu et al. [8], is a distributed clustering protocol to support both energy-homogeneous and energy-heterogeneous networks. To form clusters of unequal sizes, EADUC assigns different radius of competition to network nodes, so that CHs closest to the BS have smaller cluster sizes, thus conserving energy for inter-cluster data transfer. The protocol operates in two main phases. The first phase, known as the installation phase, consists of three sub-phases: information gathering from neighboring nodes, competition between CHs, and cluster formation. During information gathering, nodes broadcast messages containing their identifier and residual energy, enabling the construction of neighbor's tables and the calculation of average residual energy. CH competition is based on a waiting time calculated individually by each node, and the CHs formed broadcast messages for cluster formation. The second phase, data transmission, involves direct intra-cluster communication and inter-cluster data aggregation by the CHs. However, the radius of competition is determined solely on the basis of distance from the BS, taking no account of energy criteria or the number of neighbors. As a result, load balancing between CHs is not fully guaranteed.

The UCR [9] (Unequal Cluster-based Routing) protocol proposes to form clusters of unequal size to alleviate the hotspot problem. Initially, it selects CH nodes according to a random threshold, with CHs close to the BS having smaller clusters. Cluster formation is then achieved by broadcasting CH messages, and normal nodes join the cluster of the nearest CH. As far as routing is concerned, the protocol introduces a distance threshold, allowing CHs close to the BS to send data directly, otherwise they select other CHs for relaying. Although UCR balances energy consumption and improves network lifetime compared with other protocols, its main drawback is the high generation of control messages, resulting in high energy consumption and an unfair load on the CHs.

EECS [10] (Energy Efficient Clustering Scheme) is a clustering protocol for wireless sensor networks, proposed by Ye et al. It aims to balance intra-cluster energy consumption and inter-cluster communication load. CH-candidates compete by broadcasting their residual energy, and a node becomes CH if it cannot find neighboring nodes with higher energy. Unlike LEACH, EECS chooses the CH by considering both the energy of the node and the load balancing of the CHs. It uses a cost function based on the distance between the node and the CH, as well as the distance between the CH and the BS, with a weighting factor. Although EECS balances intra-cluster energy consumption and mitigates hotspots, it has limitations, such as long-distance transmissions that can deplete CH energy, making it less suitable for large-scale networks. In addition, it requires global knowledge of the distances between the CHs and the BS.

3 THE ENERGY MODEL AND THE OPTIMUM NUMBER OF CLUSTERS

3.1 Energy consumption model

The energy consumption model proposed by Heinzelman et al. [11] includes considerations for various energy-consuming activities. These include: (i) Transmission energy ($ETx(l, d)$) as defined by Formula 1, which accounts for the energy required for transmitting a message of size l over a distance d. (ii) Reception energy, denoted as $ERx(l, d)$ in Formula 3, which is calculated based on the size of the received message l. (iii) Data aggregation energy (EDA), as specified in Table 1.

$$E_{Tx}(l,d) = \begin{cases} E_{\text{elec}} * l + E_{fs} * l * d^2 & \text{si } d < d_0 \\ E_{\text{elec}} * l + E_{mp} * l * d^4 & \text{si } d \geq d_0. \end{cases} \quad (1)$$

$$d_0 = \sqrt{\frac{E_{fs}}{E_{mp}}} \quad (2)$$

E_{elec}, E_{mp}, E_{fs} are defined in Table 1.

$$E_{Rx}(l,d) = E_{Rx-elec}(l) = E_{elec} * l \quad (3)$$

Table 1: Energy model parameters.

Parameters Values	Meaning
E_{elec} = 50 nJ /bit	Required energy to run electronic circuit
E_{mp} = 0.0013 pJ /bit/ m^4	Multi path propagation
E_{fs} = 10 pJ /bit/ m^2	Free space propagation
E_{DA} = 5 nJ /bit/ signal	Required energy for Data aggregation

3.2 The optimum number of clusters K_{opt}

The K_{opt} (Optimum Number of Clusters) given by Heinzelman et al. [11] and presented in Formula 4, is determined based on various network parameters and radio characteristics (refer to Table 1). The detailed derivation can be found in [11]. In this context, M*M m^2 represents the monitoring area, K denotes the number of clusters, d_{toCH} signifies the average distance of the CHs from the BS and N is indicates the initial number of nodes.

$$K_{opt} = \frac{\sqrt{N}}{\sqrt{2\Pi}} \sqrt{\frac{E_{fs}}{E_{mp}}} \frac{M}{d_{toCH}^2} \quad (4)$$

4 ARCHITECTURE OF THE DCOPA PROTOCOL

DCOPA [4] is a distributed protocol engineered to achieve equitable clustering through a competition among network nodes. This competition is initiated by computing a T(i) value (refer to Formula 5), regarded as a timer, based on the remaining energy and the Distance to the Base Station (DistBS) for the autonomous selection of CHs in the ongoing round. $T(i) \in]0, \tau - \delta[$, less than the time allocated to the election

period of the *CHs* which is *(τ)*. $\delta \in]0, 0.1[$, is a brief duration intended to prevent a node from declaring itself as a CH outside the CH designation period. Two phases constitute DCOPA. During the setup phase (see Algorithm 1 [4]), each node decrements its own T(i) at the beginning of each round. If T(i) reaches zero, the node declares itself as a CH and broadcasts an ADV_CH message over a Radius of Clustering (RC) to its neighbors. Subsequently, these neighbors withdraw their candidacies for the CH position and await further invitations from other CHs. The steady-state phase (see Algorithm 2 [4]) comprises three periods: (i) Normal nodes transmit acknowledgment control messages to the nearest CH, (ii) CHs disseminate a Time Division Multiple Access (TDMA) schedule for sending data messages, and (iii) CHs aggregate the data and relay it to the BS using the Carrier-Sense Multiple Access Media Access Control (CSMA MAC) protocol.
Table 2 provides a description of the variables utilized in defining T(i).

Table 2: Variables of *T(i)*.

Variable	Meaning
d_{itoBS}	The distance between the node $N(i)$ and the BS.
$d_{MaxtoBS}$	The maximum distance to the *BS*.
$d_{MintoBS}$	The minimum distance to *BS*.
E_{Max}	The initial energy of the node.
E_{ri}	The residual energy of node N (i).
α	The weight of the energy criterion.
β	The weight of the distance criterion.
τ	The time of the self-election period of CHs.
δ	A small positive real number.

$$T(i) = \begin{cases} (\alpha E_i + \beta D_i)(\tau - \delta) & if\, i \in G \\ \tau - \delta & otherwise. \end{cases} \quad (5)$$

$$\alpha + \beta = 1 \quad (6)$$

$$E_i = \left(\frac{E_{Max} - Er_i}{E_{Max}}\right) \quad (7)$$

$$D_i = \left(\frac{d_{itoBS} - d_{MintoBS}}{d_{MaxtoBS} - d_{MintoBS}}\right) \quad (8)$$

E_i given in Formula 7 and D_i given in Formula 8, are defined as follow after the normalization process.
G represents the set of nodes that did not act as CHs during the previous $(1/P)$ rounds, where $P = (K/\, Nbr_{init})$. Here, Nbr_{init} stands for the initial number of nodes.

Algorithm 1 : Set up Phase

Input

$N(i).Status =' N'$,
$N(i).ClstrMbr = 0$,
$N(i).ListCH = \emptyset$,

Output

$N(i).Status$,
$N(i).ListCH$,
Compute $T(i)$,
while τ not expired **do**
  if $((T(i)$ expired) and $(N(i).ClstrMbr == 0))$ **then**
    $N(i).Status =' CH'$,
    $N(i).ClstrMbr = 1$,
    Broadcast a CtrlMsgRqJoinCH on RC,
  end if
  if (Reception of CtrlMsgRqJoinCH from $N(j)) \wedge (N(i).ClstrMbr == 0)$ **then**
    $N(i).ClstrMbr = 1$,
    $N(i).ListCH = N(i).ListCH \cup N(j)$,
    Cancel $T(i)$,
  end if
  if (Reception of CtrlMsgRqJoinCH from $N(j)) \wedge (N(i).ClstrMbr = 1) \wedge (N(i).Status =' N')$ **then**
    $N(i).ListCH = N(i).ListCH \cup N(j)$,
  end if
end while

Algorithm 2 : Steady state Phase

Input

$N(i).NbrNdCH = 0$,
$N(i).ListNdCH = \emptyset$,
$N(i).NumCH = 0$,
$N(i).Status$,

Output

$N(i).NbrNdCH$,
$N(i).ListNdCH$,
if $(N(i).Status ==' CH')$ **then**
  while (TimeWtAckMsgNds not expered) **do**
    if (Reception of CtrlMsgAckNd from $N(j)$) **then**
      $N(i).NbrNdCH = N(i).NbrNdCH + 1$,
      $N(i).ListNdCH = N(i).ListNdCH \cup N(j)$,
    end if
  end while
  Send TDMA Schdule for $N(i).ListNdCH$,
else
  $N(i).NumCH = j\ /\{N(j) \in N(i).ListCH$ is the nearest CH from $N(i)\}$,
  Send CtrlMsgAckNd for $N(N(i).NumCH)$,
  receve TDMA schedule from $N(N(i).NumCH)$,
end if
if $(N(i).Statut ==' CH')$ **then**
  while (TimeWtDataMsgNds not expered) **do**
    receve DataMsgNd (All nodes of $N(i).ListNdCH$),
  end while
  Data Agregation,
  Send MsgDataAg to the BS,
else
  Send DataMsgNd to the Node $N(N(i).NumCH)$,
end if

Algorithm 1 [4] outlines the steps of the Setup phase. The parameters $d_{MintoBS}$, $d_{MaxtoBS}$ and RC are calculated at the beginning of the first round. The variables used in this phase are described below.

N(i)	Node number i.
N(i).Status	Node's status (CH or N).
N(i).ClstrMbr	Equal to 1 if the node is part of a cluster or 0 otherwise
N(i).ListCH	List (set) of CHs requesting a normal node.
CtrlMsgRqJoinCH	Control Message for Requesting to Join a Cluster Head.

Algorithm 2 [4] outlines the steps of the steady state phase. The variables used in this phase are described below.

N(i).NbrNdCH	Number of nodes joined to a given CH.
N(i).ListNdCH	List (set) of nodes joined to a given CH.
N(i).NumCH	CH number of a normal node.
TimeWtAckNds	Time for Waiting Acknowledgments (Acks) of Nodes.
CtrlMsgAckNd	Acknowledgment Control Message
TimeWtDataMsgNds	Time for Waiting Data Messages of the Nodes.
DataMsgNd	Node Data Message.
DataMsgAg	Aggregated Data Message
TimeSDataCHtoBS	Time for Sending Cluster Head Data Message to the BS.

4.1 Calculation of the RC in DCOPA

The RC calculation of a CH is a key aspect of the DCOPA protocol, in order to optimize the announcement distance of its status as a CH. This means finding the distance over which all other candidates for the CH role will be removed, and also partitioning the network into balanced clusters. The radius R of a cluster, considering K as the optimal number of clusters and M*M m2 as the area of the monitored zone, is given by $R = \frac{M}{\sqrt{\Pi K}}$, Clusters are shaped like a circle. If we take the same conditions as in [11], a 100m*100m square, (x=50m, y=175m) for the SB position and the same radio parameters (see Table 1), we obtain K=5, in which case the radius of each circle is R ≃ 25m. Using Formula 4, we obtain K=5, in which case the radius of each circle is R ≃ 25m, making four circles that do not exceed the square of the monitored area, with no intersection. The distance between two CHs in this case is 2R ≃ 50 m, which is not appropriate in our case, as the goal is not to find the number of circles of radius R in the monitored area, which are distinct without intersection and must not extend beyond the monitored area. Circles can overflow the network surface with intersection zones and avoid having nodes without clusters.

A CH node eliminates the candidacies of other nodes on a radius $R = \frac{M}{\sqrt{\Pi K}}$, one or several nodes may be CHs located at the outer limit of the circle of radius R, in which case, if RC = $R = \frac{M}{\sqrt{\Pi K}}$, the minimum distance between any CHs is (R+ ε). We therefore won't have a number of clusters close to K, but a great number of CHs, up to twice as many (i.e. 2K), because the distance between two CHs, which must be 2R, is not respected. In consequence, for two CHs to be separated by a minimum distance of 2R, RC= $2R = \frac{2M}{\sqrt{\Pi K}}$ is required, as explained with diagrams in [4], which means in this context that the minimum distance between two CHs is (2R+ ε), in which case we can obtain a number of clusters close to K.

5 ARCHITECTURE OF THE UDCOPA PROTOCOL

5.1 Introducing ARC for unequal clustering

DCOPA belongs to the class of usual clustering algorithms that rely on a single Radius of Clustering (RC) to organize the nodes of a network into multiple clusters. However, this approach may not be optimal as nodes may have different capabilities, for example in terms of residual energy and DistBS. Our new approach, called UDCOPA [3], allows us to define for each CHs its ARC, which will be sensitive to the local criteria of the CH in question, which are residual energy and DistSB. The ARC is modeled as a multi-criteria system applied to each CH once it has been elected.

5.2 Compute and demonstrate the ARC.

Before the ADV_CH message is broadcast by a CH, the ARC is calculated using the weighted sum, which is a method applied in multicriteria analysis, of the two criteria of residual energy and the DistBS, by assigning weights which will be defined according to the IoT application or the user's needs.

The ARC takes the maximum value equal to RC (see the Formula 10) which is calculated, specified and demonstrated in [4]. RC is calculated as a function of the surface area of the surveillance zone and the optimum number of culsters K=K_{opt} (see the Formula 4).

$$RC = \frac{2M}{\sqrt{\Pi K}} \qquad (10)$$

Optimizing the ARC requires maximizing residual energy and DistBS. Thus, clusters with a large radius will be obtained when the energy and DistBS of the CH are high. However, nodes with a low amount of energy will encounter difficulties in obtaining a significant ARC, which can lead to an increased number of CHs communicating directly with the BS as their energy decreases. This can lead to a significant loss of nodes in the network, affecting its lifetime. To mitigate this problem, a minimum value of ARC, named RC_Min, is defined in Formula 11.

$$RC\_Min = \frac{RC}{3} = \frac{2M}{3\sqrt{\Pi K}} \qquad (11)$$

We establish the ARC used by a CH_i to broadcast its ADV_CH_i message using Formula 12, which we will elaborate on in the following explanation. The parameters and weights of the criteria are explained in Table 3.

$$ARC_{CH_i} = \left(\left(\theta E_{CH_i} + \omega D_{CH_i}\right)(RC - RC\_Min)\right) + RC\_Min \qquad (12)$$

$$\theta + \omega = 1 \qquad (13)$$

$$E_{CH_i} = \left(\frac{Er_{CH_i}}{E_{Max}}\right) \qquad (14)$$

$$D_{CH_i} = \left(\frac{d_{CH_i toBS} - d_{MintoBS}}{d_{MaxtoBS} - d_{MintoBS}}\right) \qquad (15)$$

The demonstrations are provided in [3].

$$RC\_Min < ARC_{CH_i} \leq RC \quad (16)$$

Formulas 12 and 16 calculate the adjusted clustering radius ARC of a CH_i as a function of its two parameters E_{CH_i} and D_{CH_i}, the coefficients θ and ω, also known as weights, determine their respective importance. RC_Min represents the minimum clustering radius chosen (see Formula 11), and RC defined as the maximum possible clustering radius calculated in DCOPA [4]. In this way, the ARC for each CH_i lies between RC_Min and RC, allowing dynamic adaptation of the radius to the specific conditions of a noued CH_i.

The minimum value of ARC which tends towards RC is obtained when a CH_i has a residual energy very close to the initial energy E_{Max} and is also very close to the BS. Inversely, the minimum value of the ARC which is RC_Min is reached when the CH_i has a very Low residual energy and is very far from the BS.

In short, when E_{CH_i} and D_{CH_i} are low, ARC tends towards RC_Min, and when E_{CH_i} and D_{CH_i} are high, ARC tends towards RC.

Table 3: Parameters of Formula 12

Parameters	Meaning
θ	The weight of the energy criteria.
ω	The weight of the distance criteria.
E_{Max}	The initial energy of the node
Er_{CHi}	The residual energy of the CH_i
$d_{CHi\,toBS}$	The distance separating the CHi from the BS.
$d_{MaxtoBS}$	The maximum distance to the BS.
$d_{MintoBS}$	The minimum distance to the BS.

5.3 UDCOPA phases

The UDCOPA protocol [3] is composed of two phases, the setup phase and the steady state phase. In the setup phase, each node begins to decrement its T(i) (see Formula 5) at the beginning of each round. If T(i) reaches zero, the node announces that it is a CH and diffuses an ADV_CH to its neighbors located in an area covered by the ARC (see Formula 12), instead of the RC introduced in DCOPA, which is the same for all CHs throughout the lifetime of the network. The Setup phase is the same as in Algorithm 1, with the modification introduced in instruction 11, replacing the RC calculated in Formula 10 with the ARC calculated in Formula 12. Nodes that receive this ADV_CH message give up their application for the role of CH and wait for further ADV_CH requests from other CHs if they are included in their ARC. The steady state phase is exactly the same as in the DCOPA protocol, see Algorithm 2.

6 SCALABILITY: PERFORMANCE EVALUATION OF UDCOPA

6.1 Scalability study goals

In the context of IoT and WSNs, scalability is a key feature for guaranteeing the efficiency and robustness of deployed systems. Scalability refers to the ability of a protocol to maintain its performance and functionality as it expands to support larger networks or wider environments. Our objective is to study and explore the geographical and quantitative scalability of the UDCOPA protocol, which was not introduced in [3] when the protocol was first published, highlighting the importance of scalability in maintaining the robustness of the proposed solution.

6.1.1 Impact of geographic scalability

Geographical scalability refers to the efficiency of a clustering protocol over large geographical areas. Specific objectives include evaluating the performance of clustering protocols over large geographical areas, analyzing the impact of distance between nodes, CHs and BS on energy performance, and determining the geographical limits beyond which a protocol does or does not lose its effectiveness. This scalability is extremely important in IoT environments where sensor nodes can be distributed over large areas such as cities, farms or industrial zones. A geographically scalable clustering protocol ensures that the network remains effective even when the distance between its component devices increases considerably.

6.1.2 Impact of quantitative scalability

Quantitative scalability refers to the effectiveness of a clustering protocol in handling an increasing number of nodes in a network. Specific objectives include evaluating the performance of clustering protocols with an increasing number of nodes, analyzing the impact of node density on energy consumption, and determining the quantitative limits beyond which a protocol loses its efficiency. This scalability is crucial in IoT environments, where the number of nodes can vary considerably, from a few dozen to several thousand. A quantitatively scalable clustering protocol ensures that the network remains efficient even when the number of nodes increases significantly.

6.2 Scalability in extended IoT environments and massive deployments

6.2.1 UDOCPA : Quantitative scalability

This section undertakes an in-depth analysis of the quantitative scalability of the UDCOPA protocol in massive node deployment. The evaluation also examines the ability of UDCOPA to maintain energy performance in large configurations and explores the distribution of CHs and the geographical extent of the clusters formed. The objective is to understand the performance of UDCOPA in massive node deployment scenarios.

6.2.2 UDOCPA : Geographic scalability

This second section looks at the performance of UDCOPA in geographically extended environments, aligned with the questions in our research motivation. The analysis explores the behavior of UDCOPA in the face of variations in the areas monitored, focusing on the response to geographical challenges, including node mortality and network lifetime. By examining the impact of geographical extent on the mortality rate, the study also sheds light on the distribution of CHs and the geographical extent of the clusters formed. This approach contributes to understanding the performance of UDCOPA in large-scale deployments, in line with the key criteria of our study and research framework.

By evaluating how unequal clustering within UDCOPA scales to large-scale IoT networks, this study aims to provide a performance evaluation that completes the study of the UDCOPA protocol.

6.3 Parameters and simulation environment

The monitoring area is a square of side M m and area M*M m^2. N nodes are deployed randomly and uniformly. All nodes are equipped with the same initial energy. Two types of messages are used for communications: control messages, which are used to organize the network, and data messages, which contain the data collected by the nodes for the application in question. The simulations are carried out using MATLAB to evaluate the performance of the protocol. The BS, responsible for data collection and processing, is outside the monitored area. We have made further assumptions about the properties of the network and nodes. The weights of the criteria, used in T(i) and ARC_{CHi}, that we have chosen are listed in the Table 4.

6.4 Assumptions

In our simulations, we have made several assumptions about the characteristics of the BS and the network nodes. First of all, the BS has unlimited energy, which means that it can operate without any power constraints. The nodes, on the other hand, are equipped with batteries that cannot be replaced or recharged. In addition, nodes are not mobile and do not have the technological equipment needed to know their positions. Finally, nodes have the ability to adjust their transmission range according to their distance from the receiver(s) and will fail if and only if their energy is completely depleted.

Table 4: Simulation parameters

Parameters	Meaning
$Sink_x$ = X m	sink x-axis
$Sink_y$ = Y m	sink y-axis.
$MsgCtrl$ = 25 $bytes$	Control Message length.
$DataMsg$ = 200 $bytes$	Data Message length.
K = K_{opt} $clusters$	Optimum clusters number [11].
$M * M$ m^2	Area network

Parameters	Meaning
$EMax = 0.5\ j$	Initial energy

6.5 Mortality rate and lifetime parameters

The lifetime parameters we have chosen are as follows: FND (First Node Die), QND (Quarter Nodes Die), HND (Half Nodes Die), SND (Seventy-five Percent Nodes Die), and LND (Last Node Die). Simulations, performed by varying the number of nodes from 100 to 1000 with an increment of 100 nodes at each iteration on a deployment surface of 200 *200 m^2, 400*400 m^2, 600 *600 m^2 and 800*800 m^2, as presented in Tables 5, 6, 7 and 8 respectively, clearly reveal the maintenance of UDCOPA's performance in terms of lifetime parameters as the number of nodes in the network increases as well as the geographical surface increases, as illustrated in Subfigures (a), (b), (c) and (d) of Figure 1.

Table 5: **Lifetime parameters by increasing the number of nodes in a 200 $*$ 200 m^2 area**

Nodes	Area (m^2)	RC	BS position (m, m)	FND	QND	HND	SND	LND
100	200^2	71	(100,275)	88	346	500	619	1152
200		60		82	427	632	793	1228
300		55		92	453	577	791	1240
400		52		173	516	667	821	1300
500		48		58	544	663	833	1235
600		46		166	564	694	841	1273
700		44		117	584	736	878	1242
800		43		100	607	738	900	1195
900		42		74	615	750	910	1235
1000		41		70	644	761	918	1192

Table 6: **Lifetime parameters by increasing the number of nodes in a 400 $*$ 400 m^2 area**

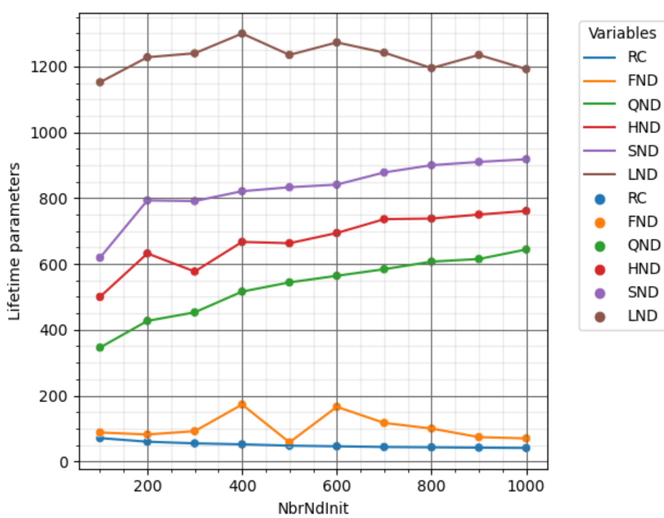
Nodes	Area (m^2)	RC	BS position (m, m)	FND	QND	HND	SND	LND
100	400^2	104	(200,475)	6	39	99	315	1045
200		87		2	61	171	377	1158
300		77		2	66	182	335	1389
400		72		2	93	236	427	1123
500		69		3	96	234	472	1465
600		66		2	110	280	497	1285
700		63		2	119	296	497	1286
800		61		2	118	250	434	1343
900		59		3	122	265	491	1277
1000		58		2	130	307	515	1218

Table 7: **Lifetime parameters by increasing the number of nodes in a 600 $*$ 600 m^2 area**

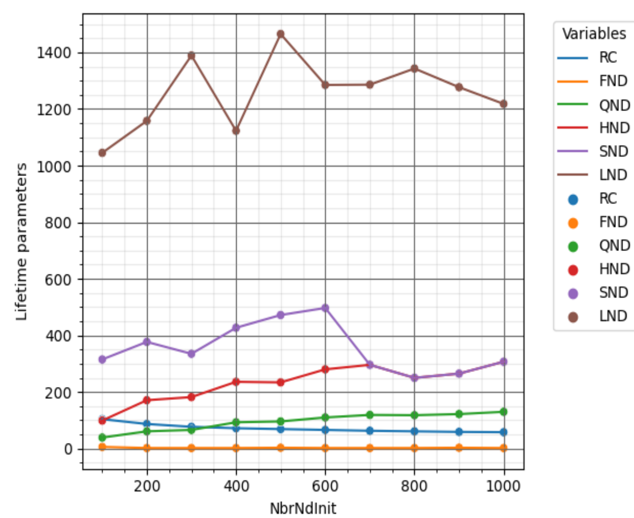
Nodes	Area (m^2)	RC	BS position (m, m)	FND	QND	HND	SND	LND
100	600^2	126	(300,675)	1	11	25	63	635
200		106		1	17	43	131	790
300		96		1	23	54	147	812
400		89		1	25	53	150	788
500		84		1	26	57	133	980
600		80		1	29	69	168	1361
700		77		1	32	72	180	1248
800		75		1	36	85	244	1354
900		73		1	36	80	218	1238
1000		71		1	38	80	225	1487

Table 8: **Lifetime parameters by increasing the number of nodes in a 800 ∗ 800 m^2 area**

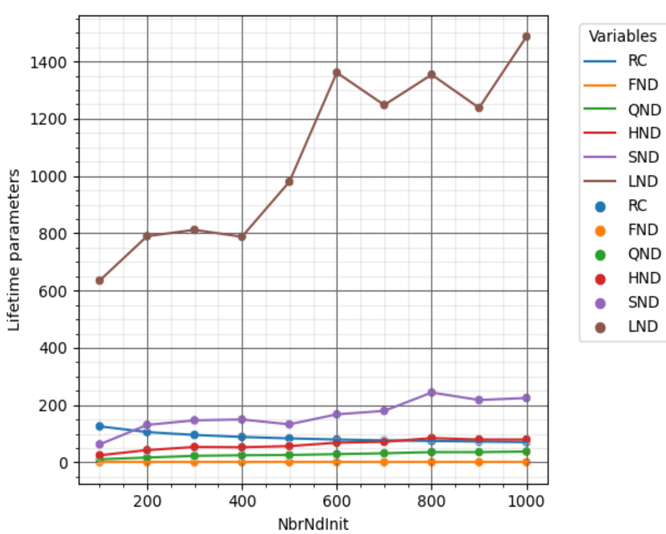
Nodes	Area (m^2)	RC	BS position (m, m)	FND	QND	HND	SND	LND
100	800^2	145	(400,875)	1	9	15	32	332
200		122		1	10	23	41	930
300		110		1	13	28	60	1185
400		102		1	15	32	63	1072
500		97		1	18	38	78	1315
600		93		1	18	39	80	1205
700		89		1	20	41	90	1230
800		86		1	21	44	98	1257
900		84		1	23	48	105	1430
1000		81		1	24	49	101	1490



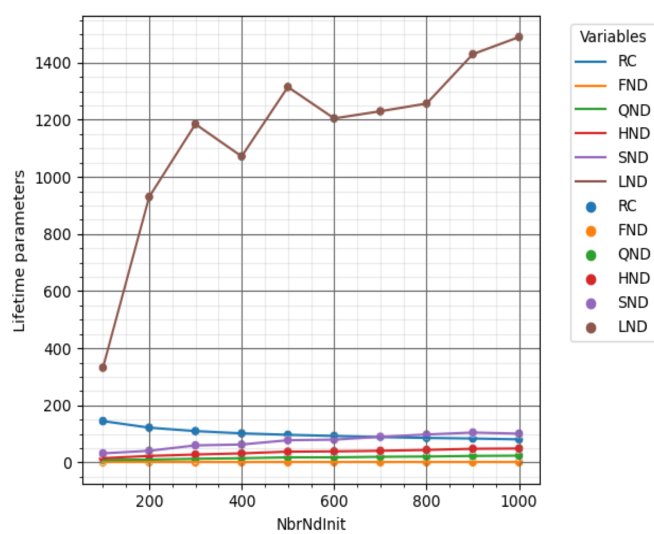
(a) Area 200 m^2



(b) Area 400 m^2

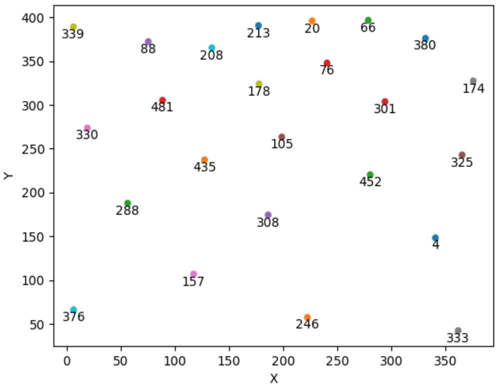


(c) Area 600 m^2

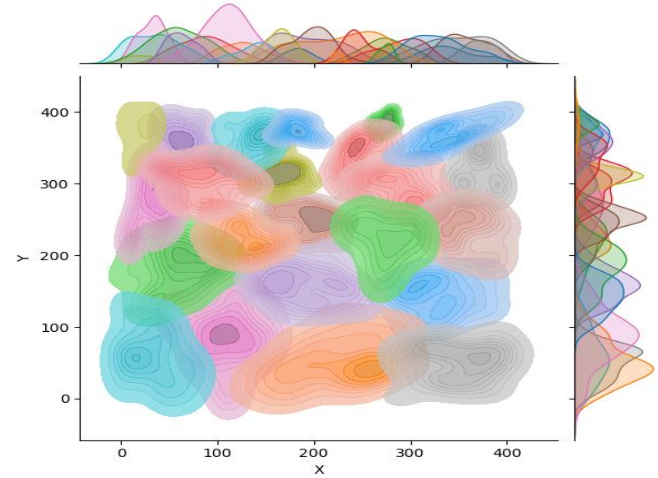


(d) Area 800 m^2

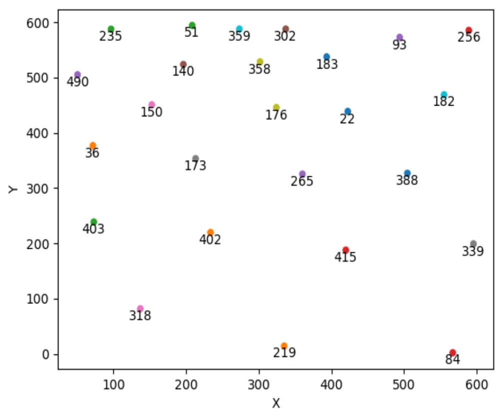
Figure 1: Lifetime parameters by increasing the number of nodes in different areas



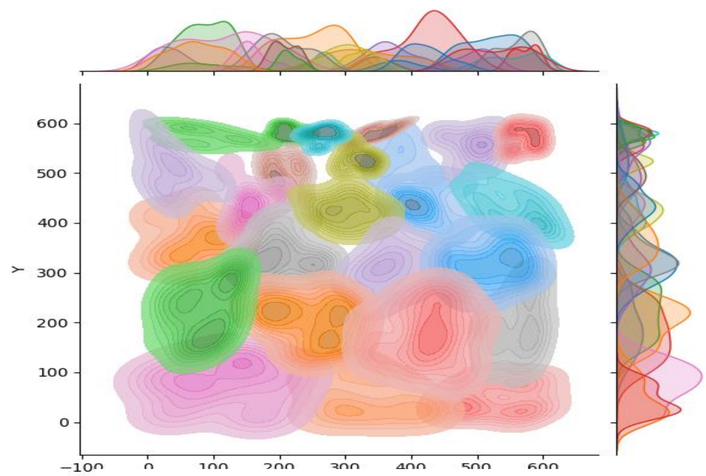
(a) Distribution of CHs (Area 400 m^2)



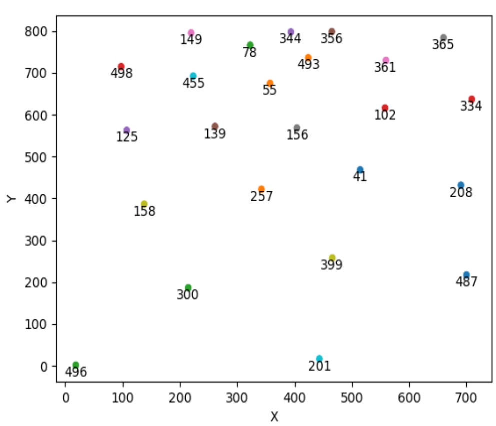
(b) Distribution of clusters (Area 400 m^2)



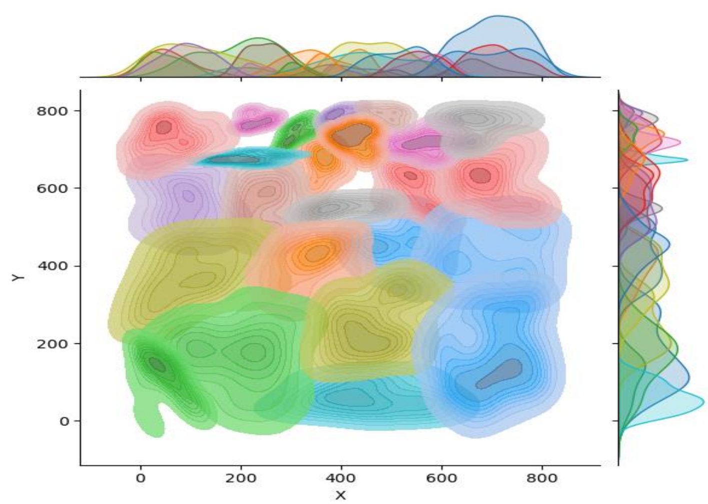
(c) Distribution of CHs (Area 600 m^2)



(d) Distribution of clusters (Area 600 m^2)



(e) Distribution of CHs (Area 800 m^2)



(f) Distribution of clusters (Area 800 m^2)

Figure 2: Distribution of CHs and clusters with 500 nodes and various areas

6.6 Distribution of CHs and clusters in the Network

A balanced distribution of CHs in a clustering protocol is a key aspect in optimizing network performance. Ensuring a balanced distribution of CHs over the monitored area guarantees that the processing and communication load is evenly distributed, preventing the over-utilization of certain nodes and thus prolonging node and network lifetime. In addition, this approach reduces the premature death rate of nodes, by avoiding hot spots and rapid energy depletion, and improves fairness in the rotation of CH roles, enabling all nodes to participate equally in the management of network activities, especially communication.

The distribution of CHs presented in Subfigures (a), (c) and (e) of Figure 2 for surfaces of 400*400 m^2, 600*600 m^2 and 800*800 m^2, respectively, clearly shows the maintenance of a very homogeneous distribution of CHs as the deployment surface increases. We also observe the same trend for the geographical extent of the clusters, illustrated in sub-figures (b), (d) and (f) of Figure 2 for surfaces of 400*400 m^2, 600*600 m^2 and 800*800 m^2, respectively. These clusters are well distributed across the deployment area, ranging from large clusters located far from the BS, to medium-sized clusters in intermediate areas of the surface, to small clusters in areas very close to the BS.

6.7 Discussion on UDCOPA scalability: quantitative and geographical.

Increasing the deployment area or the number of nodes did not influence the network lifetime parameters or the mortality rate, or even the balanced distribution of CHs across the network and the geographical extent of the clusters built. This is due to the adaptation of the clustering radius of the CHs, calculated as a function of their energy and their distance from the BS. The K_{opt} number plays a key role in this configuration, as it increases with the number of nodes and the monitoring area. Similarly, the base clustering radius, on which UDCOPA relies to reduce the radius, is calculated as a function of the deployment area in order to position the K_{opt} clusters.

7 CONCLUSION

The UDCOPA [3] protocol is a significant improvement on the DCOPA [4] protocol for unequal clustering in data communications within IoT networks. The main objective of UDCOPA has been to optimize the adaptability of the clustering radius, the sensitivity to CH contextual criteria and the dynamic adjustment of the clustering radius to balance energy consumption, cluster extent and the number of nodes per cluster. The aim of this research work is to test the geographical and quantitative scalability of the UDCOPA protocol, which was not included when it was first presented by the authors. The study of scalability was at the heart of this research work, with various simulations covering a wide range of configurations. By adjusting the number of nodes deployed in a fixed geographical area, we were able to evaluate UDCOPA's performance in different contexts, namely mortality rate, distribution of CHs and clusters, and geographical extent of clusters. The results of our simulations revealed that UDCOPA is scalable in terms of energy management, node and network lifetime, and distribution of CHs and clusters in the network, maintaining the same performance despite changes in the space monitored and the number of nodes in the network.

REFERENCES

[1] OUBELAID, Adel, ALBALAWI, Fahad, REKIOUA, Toufik, et al. Intelligent torque allocation based coordinated switching strategy for comfort enhancement of hybrid electric vehicles. IEEE Access, 2022, vol. 10, p. 58097-58115.

[2] OUBELAID, Adel, KHOSRAVI, Nima, BELKHIER, Youcef, et al. Health-conscious energy management strategy for battery/fuel cell electric vehicles considering power sources dynamics. Journal of Energy Storage, 2023, vol. 68, p. 107676.

[3] MIR, Foudil et MEZIANE, Farid. Unequal Clustering Protocol in IoT Networks Based on Multiple Criteria Processing. In : Proceedings of the 2023 6th International Conference on Information Science and Systems. 2023. p. 132-136.

[4] MIR, Foudil et MEZIANE, Farid. DCOPA: a distributed clustering based on objects performances aggregation for hierarchical communications in IoT applications. Cluster Computing, 2023, vol. 26, no 2, p. 1077-1098.

[5] MIR, Foudil et MEZIANE, Farid. Novel adaptive DCOPA using dynamic weighting for vector of performances indicators optimization of IoT networks. Expert Systems with Applications, 2024, vol. 247, p. 123212.

[6] MALATHI, L., GNANAMURTHY, R. K., et CHANDRASEKARAN, Krishnan. Energy efficient data collection through hybrid unequal clustering for wireless sensor networks. Computers & Electrical Engineering, 2015, vol. 48, p. 358-370.

[7] LIAO, Ying, QI, Huan, et LI, Weiqun. Load-balanced clustering algorithm with distributed self-organization for wireless sensor networks. IEEE sensors journal, 2012, vol. 13, no 5, p. 1498-1506.

[8] YU, Jiguo, QI, Yingying, WANG, Gang, et al. An energy-aware distributed unequal clustering protocol for wireless sensor networks. International Journal of Distributed Sensor Networks, 2011, vol. 7, no 1, p. 202145.

[9] CHEN, Guihai, LI, Chengfa, YE, Mao, et al. An unequal cluster-based routing protocol in wireless sensor networks. Wireless Networks, 2009, vol. 15, p. 193-207.

[10] YE, Mao, LI, Chengfa, CHEN, Guihai, et al. EECS: an energy efficient clustering scheme in wireless sensor networks. In : PCCC 2005. 24th IEEE International Performance, Computing, and Communications Conference, 2005. IEEE, 2005. p. 535-540.

[11] HEINZELMAN, Wendi B., CHANDRAKASAN, Anantha P., et BALAKRISHNAN, Hari. An application-specific protocol architecture for wireless microsensor networks. IEEE Transactions on wireless communications, 2002, vol. 1, no 4, p. 660-670.