Novel Adaptive DCOPA Using Dynamic Weighting for Vector of Performances Indicators Optimization of IoT Networks

Foudil Mir[a,*], Farid Meziane[b]

[a] *Medical Computing Laboratory (LIMED), Faculty of Exact Sciences, University of Bejaia, 06000 Bejaia, Algeria*
[b] *Data Science Research Centre, University of Derby, UK*

Abstract

Clustering in the Internet of Things (*IoT*) involves organizing devices into groups to streamline network management and optimize resource utilization, including Internet connections, energy usage, coverage, quality of service, and connectivity. *DCOPA* (A Distributed Clustering Based on Objects Performances Aggregation for Hierarchical Communications in *IoT* Applications) is a recent distributed clustering protocol based on a timer for cluster formation where the election of Cluster Heads (*CHs*) is modeled as a multicriteria problem. In this paper, three contributions are presented. Firstly, the *DCOPA* protocol is analyzed with a focus on its multi-criteria aggregation function *T(i)* which directly contributes to the election of the *CHs* and the formation of the network's clusters. This is then followed by an in-depth analysis of the impact and the variation of the weights assigned to the two aggregated criteria which are the energy and the distance from the base station. A verification of the scalability, load balancing and distribution of the clusters and *CHs* will follow. Secondly, a new formal notation for the performance analysis, specifically focusing on the mortality and lifetime based on the Vector of Performance Indicators (*VPI*), will be introduced for *IoT*. As a third contribution, a revised version of *DCOPA*

*Corresponding author
Email addresses: `foudil.mir@univ-bejaia.dz` (Foudil Mir), `f.meziane@derby.ac.uk` (Farid Meziane)

is introduced called *ADCOPA* (Adaptive *DCOPA* Using Dynamic Weighting for Vector of Performances Indicators Optimization of *IoT* Networks). *ADCOPA* is based on a new property which is the dynamicity or the variability of the weights of the criteria used in the election function of *CHs*. The simulation results show that the *ADCOPA* algorithm, which dynamically adjusts the weights of the criteria during the network's lifetime, outperforms the *DCOPA* algorithm. The latter uses static weights for the criteria that remain unchanged for the entire lifetime of the network. This confirms that the ability to dynamically adjust the weighting of the criteria is an important factor in achieving better performance.
Keywords: *IoT*, *DCOPA*, clustering, data communication, energy efficiency, multi-criteria analysis, dynamic weights of criteria, *VPI*.

1. Introduction, research questions and motivations

1.1. Introduction

The Internet of the future contains billions of communicating smart "things" of which the *IoT* is a part of (Li et al., 2015). The global reach of *IoT* is immense and increasing exponentially (Nord et al., 2019). The advancement of *IoT* is achieved through the development of several technologies such as Radio-Frequency IDentification (RFID) and sensor networks (Al-Fuqaha et al., 2015). The vast majority of electronic devices used are sensors derived from recent advances in Micro-Electro-Mechanical Systems (MEMS) (Akyildiz et al., 2002). In a sensor network application dedicated to *IoT*, the data collected by a node is sent to a processing station called Base Station (*BS*). The research work on the functioning of communication protocols with energy resource optimization for this type of network is still relevant. *IoT* is virtually organized in several groups called "clusters". Each cluster has a specific node or coordinator called Cluster Head (*CH*) which is used to aggregate the data received from the nodes of the same cluster and then communicate it to the *BS*. This division into clusters is called "clustering". Energy awareness is one of the main concerns in recent advances in wireless sensor networks (WSNs) for *IoT*. Clustering is the most commonly used strategy in data communication to achieve energy efficiency in WSNs frequently deployed in hostile environments, often characterized by inaccessibility. Its main objective is to extend the life of networks and keep their applications operational. In this type of environments, where battery replacement may be impossible, clustering optimizes energy usage by minimizing data transmissions and the number of Internet connections in the case of IoT applications, thereby increasing the battery life of equipment and sensors, enabling them to continue operating for long periods. This approach is very useful in hostile environments such as geographically isolated regions, deserts or high-risk security locations. It is used, for example, to monitor rare species, toxic environments, critical infrastructures and to collect measurements in difficult circumstances. The clustering technique enables surveillance to be extended, optimizing energy consumption in particular. Due to the selection of inappropriate *CHs*, long distance data transmissions have a negative effect on the network efficiency in terms of stable period, lifetime and quality of service (QoS) (Benelhouri et al., 2023). Dedicated clustering for data transmission optimizes energy consumption in WSNs applications, which is an expanding field especially with the emergence of IoT. The choice of nodes that will take the role of *CHs* is a complex task that must take into account several challenges. There are two types of approaches for selecting the *CHs*: distributed as illustrated in (Heinzelman et al., 2000) and centralized as described in (Heinzelman et al., 2002). Furthermore, it is necessary to define the algorithm that rotates the nodes of the network in the role of *CH*, as well as the criteria determining this election. The selection of these nodes must be based on a careful analysis of several parameters and factors while ensuring optimal network operation. This process must also take into consideration the efficient and balanced energy management of the nodes and the entire network to guarantee a better lifetime of the nodes and the entire network. LEACH is one of the first dedicated probabilistic and distributed clustering protocols for hierarchical routing that proceeds in two main phases. *DCOPA* (Mir & Meziane, 2023) is a distributed clustering algorithm based on Multi-Criteria Decision Making (MCDM) modeling (Triantaphyllou, 2000). The election of the *CHs* is achieved through the use of a competition where all the sensors are engaged based on a timer *T(i)* computed according to two local criteria, the residual energy of the node and its Distance from the Base Station (*DistBS*). *T(i)* is computed based on multi-criteria aggregation with predefined weights that are associated to the two chosen criteria. *DCOPA* considerably improves the deficiencies present in the LEACH protocol, namely the distribution of the *CHs*, the announcement of a *CH* which is done by the sending of a message on a maximum distance which covers all the network and the number of *CH* which is purely random. The aim of our work is to analyze and enhance the existing *DCOPA* protocol further in such a way that it increases the network lifetime with balanced clustering and better energy efficiency. *DCOPA* incorporates the energy and the *DistBS* of a node as two main criteria that determine its eligibility to win the role of the *CH* by calculating a *T(i)* which is a weighted sum of these two parameters. Each criterion is assigned a specific weight that reflects its relative importance compared to the other criteria. This weighting allows to prioritize certain criteria over others when evaluating alternatives.

1.2. Research questions

This study is guided by a number of research questions focusing on (i) the analysis of the multi-criteria aggregation function of the DCOPA protocol (recently proposed in the field of clustering with energy optimization in *IoT* applications); (ii) the impact of dynamic weighting on the criteria used (energy and distance); (iii) and the design of a new mechanism for analyzing *IoT* performance, in particular energy management and mortality rates. These research questions are formulated as follows:

1. What is the impact of applying different combinations of static weights (maintained until the loss of all nodes) when applied to the multi-criteria aggregation function of the *DCOPA* protocol on the *CHs* election process, cluster formation and network energy management?

2. What impact will the different combinations of static weights have on scalability, energy load balancing and the distribution of clusters and *CHs*?
3. To what extent do the weight of the energy criterion and the distance to the base station influence the *CH* designation process?
4. To what extent do the performance parameters mentioned above improve with the incorporation of dynamic weighting in the *DCOPA* protocol?
5. What are the key factors and metrics that indicate when the weights can be modified to contribute to a better performance, such as energy optimization and mortality rates?
6. What new mechanism could be designed to effectively analyze *IoT* performance based on mortality rate parameters (lifetime)?

To provide systematic answers to these research questions, several simulation cases of the *DCOPA* protocol were performed by varying the (α, β) combinations as well as the number of nodes in the network in order to evaluate the parameters of lifetime, scalability, load balancing and distribution of *CHs* and clusters. A formal notation is introduced to evaluate and compare two protocols in terms of network degradation or node failure rate. After an evaluation of *DCOPA*, an improvement is proposed. The *BS* has the privilege to influence the function *T(i)* by sending to the nodes the combination of the weights of the parameters (criteria). The change of the weights of the criteria is operated according to the data and the state of the network which was communicated to it through the *CHs* nodes of each round. This implies that the dynamicity of the weights has been introduced in the multicriteria evaluation, precisely in the weighted sum. Therefore, a revised version of *DCOPA* is introduced named *ADCOPA*, adapting the principle of dynamicity of the weights of the criteria and context awareness regarding the mortality identified by the *BS*, which then decides to change the weights of each criterion in an attempt to slow down the flow of node failures. For this reason, only one change of the criteria weights is performed during the lifetime of the network, in order to illustrate the interest of the dynamicity of the criteria weights that will be operational during a specific round number that was considered significant. An analysis was carried out to determine the best combination of weights for the two aggregated criteria. In other words, a process was undertaken to identify the criterion that has the greatest impact on the *VPI* parameters, assigning it more importance (weights) before or after a chosen round number. The performance analysis of *ADCOPA*, in this case, shows its efficiency by comparing it to *DCOPA* on the level of mortality rate and lifetime considered parameters.

1.3. Research motivations

The main motivations of the current work is to improve energy management in WSNs which remains a challenging issue to improve their life and develop new protocols to extend their lives and operations. Specifically, the motivation of the current work can be summarized as follows.

- It is often more beneficial to improve existing solutions for energy-efficient clustering for *IoT* networks than designing an entirely new one. In our case, we examine and improve the *DCOPA* protocol by leveraging its strengths and addressing its weaknesses. The goal is to optimize the clustering process by incorporating a new property that enhances efficiency in terms of energy management and *CHs* selection.

- Understanding the mortality and lifetime of a data communication clustering protocol in an *IoT* network is crucial for its improvement. The introduction of a new formal notation for performance analysis, using a set of performance indicators, can provide a valuable tool to evaluate the energy efficiency and mortality rate of a clustering protocol in an *IoT* network. This allows us to better understand the strengths and weaknesses of a protocol and to make improvements to optimize its performance, especially in terms of energy efficiency and lifetime extension.

- Multi-criteria analysis is a powerful tool for decision making when several parameters need to be optimized simultaneously. The *DCOPA* protocol uses multi-criteria aggregation to select the *CHs* in each round by assigning specific weights to different criteria. However, the changing state of the network may require adaptation to improve its energy efficiency and preserve a maximum number of nodes for an extended period of time. This led to the introduction of a new aspect based on dynamic criterion weights sensitive to the energy context and the mortality of the network.

The paper is organized as follows. Related works are described in section 2. A brief overview of the *DCOPA* protocol and the energy model used are described in section 3. Section 4 is reserved for the simulation and analysis of the influence of the energy, the distance to the base station *BS* in the performances of *DCOPA*. In section 5, a new formal notation named *VPI* is introduced for evaluating the performances of a network in terms of node mortality rate. Section 6 is dedicated to the description of the *ADCOPA* protocol and its performance evaluation. The last section is dedicated to the conclusion and the future perspectives of our approach.

2. Related works

The authors in (Hosseinzadeh et al., 2022) conducted a qualitative study of the clustering algorithms in IoT precisely in the field of smart cities following a systematic literature review published between 2017 and 2021. They showed through their work the usefulness of clustering in IoT especially for energy efficiency, scalability, robustness, mobility, and load balancing. This section is dedicated to the presentation of some protocols and approaches used in the literature on clustering in IoT networks mainly based on WSNs. (Heinzelman et al., 2000) proposed LEACH, a distributed protocol for dynamic and probabilistic clustering. A threshold *T(i)* is computed by each node of the network based on a chosen percentage of *CHs* and the number of the current round. Afterwards, *T(i)* is compared to a random number between $[0, 1]$. LEACH runs in two phases, set-up and steady-state. A node $N(i)$ calculates the value *T(i)* in the set up phase. If the random number is less than *T(i)* then the node $N(i)$ declares itself as CH and sends an advertisement message (ADV-CH) to the whole network. The steady state phase is reserved for cluster formation and data communication. Ordinary nodes solicited with ADV-CH messages will choose the nearest CH by sending a JOIN-CH message. A Time Division Multiple Access (TDMA) schedule will be planned and broadcasted for intra-cluster communication to avoid collisions. The CHs aggregate the received data and send it to the BS. LEACH has attracted other researchers to improve it for better results. This is due to its interesting properties, namely the simplicity of its algorithm, the probabilistic and distributed character, the rotation of the role of the CH in a balanced way between the different nodes of the network. (Hani & Ijjeh, 2013; Rahayu et al., 2014; Arora et al., 2016; Singh et al., 2017) are examples of surveys in the literature, highlighting their significant contributions in collecting and discussing improvements of the LEACH protocol. (Heinzelman et al., 2002) improved LEACH by proposing LEACH-C, where the BS is responsible for the election of the CHs and the formation of the clusters by knowing the positions and the energy level of the nodes at the beginning of each round. The simulated annealing algorithm (Murata & Ishibuchi, 1994) is used to find the optimum number of clusters which is considered as an NP-Complete problem. (Junping et al., 2008) proposed the distributed Time Based LEACH (TB-LEACH) protocol to improve the CHs selection procedure proposed in the LEACH protocol. The authors specified the number of CHs required. At the beginning of each round, a node creates a random number that is considered as a time. If a node's time expires before the number of CHs in the network is reached, it declares itself CH. Otherwise, the node decides to leave the competition. After the election of the CHs, the process is identical to that of LEACH. The optimal number of clusters is a predefined value (K_{opt}) in the LEACH protocol, but this value can vary randomly during the execution of the clustering algorithm. This is one of the major drawbacks of the protocol. (Batra & Kant, 2016) introduced LEACH-MAC (a new cluster head selection algorithm for WSNs) describing a new strategy to make the number of CHs and clusters stable by using MAC layer information to control the randomness used in the LEACH clustering process.

The main idea of this approach is to limit the number of ADV (Advertisement) messages from the CHs. A variable named CHheard is initialized to 0 when the selection of the *CHs* is launched. It is incremented by 1 when an ADV message is received. A random time between [0, $total\_adv\_time$] is chosen by the node, where ($total\_adv\_time$) is the transmission and reception time for the *CHs*. The chosen time is (R_t), then the time to send the announcement t_{adv-CH} can be evaluated as (t_{adv-CH}=($R_t/Current\_Energy$)). At the time t_{adv-CH}, the node examines the variable CHheard to determine the number of ADV messages received, if the value of CHheard is lower than the optimum number of clusters, it declares itself *CH* and sends an ADV message. (t_{adv-CH}) ensures that nodes with more energy can send ADV messages before those with less. Maintaining a constant number of *CHs* and cluster balancing is the focus of LEACH-Balanced (LEACH-B) (Tong & Tang, 2010). The *CHs* of the LEACH set up phase broadcast a message for the whole network with their residual energies. If the number of *CHs* is greater than ($N * P$), where N is the number of nodes and P the percentage of desired *CHs*, the *CHs* that do not contain enough energy will become normal nodes. If it is the opposite, some normal nodes will become *CHs* according to a timer calculated so that the number ($N * P$) of *CHs* is ensured. Improving the $T(n)$ formula for the election of *CHs* in LEACH is the goal of the Improved-LEACH (I-LEACH) protocol (Beiranvand et al., 2013). The authors considered new criteria which are the residual energy of the node, the *DistBS* and the number of neighbors. A hybrid approach (centralized and distributed) is adopted for the selection of *CHs* and clusters in LEACH-G proposed in (Chen et al., 2013). The authors try to remedy the disadvantages of LEACH which are: (i) the optimum number of clusters (K_{opt}, see Formula 12) of *CHs* which is not reached in the majority of rounds, (ii) the number of *CHs* which is variable, (iii) the elected *CHs* are not distributed in a balanced way in the network. R-LEACH is proposed by Behera et al. in (Behera et al., 2019) containing the same phases of LEACH. The selection of *CHs* is similar to LEACH in the first round. A new $T(n)$ will be considered from the second round. $T(n)$ is calculated based on the residual energy of the node, the old LEACH $T(n)$, the initial energy of the node and K_{opt} (defined in (Hussain & Matin, 2005)). However, the steady state phase of LEACH was maintained. (Zhao et al., 2019) in Modified LEACH (MLEACH) focuses on the optimum number of *CHs* and considers energy as a key factor for the election of *CHs*. A new $T(n)$ is given as a function of the residual energy of the node, the total residual energy of the network, the energy consumed in the previous round and the average energy consumed in the previous round and by the whole network. The authors in (Chang et al., 2022) attempted to solve two drawbacks of LEACH which are the random selection of *CHs* and the unequal cluster size. The improvement is based on the distribution density of the nodes and the allocation of the remaining nodes. The authors in (Kumar et al., 2023) introduced a novel energy-efficient clustering approach for WSNs that aims to reduce energy consumption. The approach involves selecting CHs using a threshold-based advanced LEACH (ADV-LEACH2) approach and formulating clusters using a Modified Fuzzy C-Mean (MFCM) approach. The ideal *CH* is selected based on factors such as remaining energy, number of neighbor (SNs), and *DistBS*. The proposed TEEECH approach considers these factors and uses the current energy of the SNs per round to measure optimal *CH* selection. The TEEECH balances sensor node distribution among cluster heads and maintains *CH* energy. (Mir & Meziane, 2023) presented a new distributed method called *DCOPA* for clustering with energy optimization in IoT networks. *DCOPA* uses MCDM (weighted sum) theory to elect *CHs* in each round. Section 3 provides a complete and detailed description of the *DCOPA* protocol. The authors in (Srinidhi et al., 2019) presented a comprehensive study of network optimization for IoT communication, focusing on different achievements and emerging challenges in this field. The survey covers new methods and algorithms for multi-objective questions, routing protocols, energy efficiency and security in IoT networks. They provided a literature review of recent research papers and also identified relevant parameters such as routing, energy conservation, congestion, scalability, reliability, QoS and security. Advanced network optimization techniques and their limitations are categorized, including issues such as packet transmission, network overhead reduction, network attacks, device interoperability and node mobility. In contrast to traditional recommendation algorithms, which only consider accuracy as the objective of optimization, (Cao et al., 2021) investigated the building of a novel multi-objective matrix factorization model to increase the recommendation performance, LSMaOA (large-scale many-objective optimization algorithm based on problem transformation), for personalized recommendation in IoT systems. The model considers six objectives: F1 measure (comprehensive evaluation index for precision and recall), novelty, coverage, customer satisfaction, landmark similarity and overfitting. The proposed LSMaOA enhances optimization efficiency by making the large-scale problem smaller. Experimental findings confirm that it outperformed four different algorithms. (Shreyas et al., 2021) introduced an energy-efficient optimal routing technique for IoT networks. By choosing an efficient cluster head, an optimal path from source to destination is obtained. The proposal adopts Genetic Algorithm (GA) as an optimization approach to achieve optimal routing results. The fitness function is applied to allow the selection of multi-path routing, minimizing energy consumption and enhancing network lifetime. Performance evaluation has confirmed the robustness of the algorithm, proving its performance in terms of energy consumption, end-to-end delay and number of failure nodes. EFC-ISFLA is proposed in (Kongsorot et al., 2022), a fuzzy-based clustering protocol optimized by the Improved Shuffled Frog Leaping (ISFLA) algorithm, for energy-efficient energy consumption in WSNs. EFC-ISFLA selects adequate CH nodes based on an energy threshold and optimized fuzzy inference systems (FIS). Cluster building and next hop (NH) identification are also controlled by FIS probabilities. ISFLA both optimizes the network parameters and the FIS components simultaneously, using opposition-based operators and a substitution model. Simulation analysis revealed that EFC-ISFLA is more efficient than existing protocols in terms of network lifetime, stability and packet delivery to the base station. (Abdulzahra et al., 2023) presented an Energy-Efficient Fuzzy-based Unequal Clustering with Sleep Scheduling (EFUCSS) protocol for WSN-based IoT networks. The EFUCSS protocol acquires more network lifetime and minimizes energy consumption by performing clustering, scheduling and data transmission. It adopts fuzzy logic for CH selection (Fuzzy inference is obtained by using the Mamdani technique), taking into account factors such as a node's residual energy, centrality and node-GW (GateWay) distance. The proposed protocol addresses the energy hole issue by creating unequal clusters using Fuzzy C-Means. It also introduces a CH re-selection mechanism based on an energy threshold and a node matching method to decrease the number of transmitting nodes.

3. *DCOPA*, the used energy consumption model and the optimum number of clusters

3.1. Energy consumption model

The energy dissipation model of (Heinzelman et al., 2002) is used in the simulations of the *DCOPA* protocol. It accounts for the energy required for transmission as a function of the size (l) of the message and the communication distance (d). The energy of reception is calculated just according to the size of the received message (l). The energy of data aggregation is considered by this model.

During transmission, the energy consumed is defined by $E_{Tx}(l, d)$ in Formula 3 and detailed in the Formula 5. Depending on the distance d, two power control settings and two channel models are used to define the parameter $E_{Tx-amp}(l, d)$ as follows:

- If ($d < d_0$), d_0 is given in Formula 6, the free space model channel (d^2 power loss) and the free space power amplifierE_{fs} are used.

$$E_{Tx-amp}(l, d) = E_{fs} * l * d^2 \quad \text{if } d < d_0. \tag{1}$$

- Otherwise ($d \geq d_0$), the multipath fading model channel (d^4 power loss) and the multipath power amplifier E_{mp} are used.

$$E_{Tx-amp}(l, d) = E_{mp} * l * d^4 \quad \text{if } d \geq d_0. \tag{2}$$

$$E_{Tx}(l,d) = E_{Tx-elec}(l) + E_{Tx-amp}(l,d) \tag{3}$$

$$E_{Tx-elec}(l) = E_{elec} * l \tag{4}$$

As a conclusion from Formulas 1, 2, 3 and 4, the total energy consumption of transmission can be defined using Formula 5.

$$E_{Tx}(l,d) = \begin{cases} E_{elec} * l + E_{fs} * l * d^2 & if\ d < d_0. \\ E_{elec} * l + E_{mp} * l * d^4 & if\ d \geq d_0. \end{cases} \tag{5}$$

$$d_0 = \sqrt{\frac{E_{fs}}{E_{mp}}} \tag{6}$$

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

The reception energy $E_{Rx}(l,d)$ de l bits is defined in the Formula 7.

$$E_{Rx}(l,d) = E_{Rx-elec}(l) = E_{elec} * l \tag{7}$$

From Formulas 4 and 7, it is evident that Formula 8 is verified.

$$E_{\text{Tx-elec}}(l) = E_{Rx-elec}(l) = E_{elec} * l \tag{8}$$

The electronics energy, represented by E_{elec}, is influenced by various factors including digital coding, modulation, filtering, and signal spreading (Heinzelman et al., 2002).

The nodes *CHs* aggregate the data (signals) with an energy called E_{DA}, see Table 1, to send them to the *BS*.

Table 1: Energy model parameters

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

3.2. *The optimum number of clusters K_{opt}*

The optimal number of clusters (K_{opt}) given by (Heinzelman et al., 2002), which is presented in Formula 12, is computed as a function of the energy consumption of a *CH* node (Formula 9), a non-CH node (Formula 10), the energy consumed by a cluster (Formula 11), the network parameters and the radio characteristics of the used devices (see Table 1). The essential steps are described in detail in the following. The complete formal demonstration is in (Heinzelman et al., 2002). The monitoring area is M*M m^2, K is the number of clusters, d_{toCH} is the average of the distances of the *CHs* from the *DistBS* and N is the number of initial nodes.

$$E_{CH} = lE_{elec}(\frac{N}{K} - 1) + lE_{DA}\frac{N}{K} + lE_{elec}d^4_{toBS} \tag{9}$$

$$E_{Non-CH} = lE_{elec} + lE_{fs}\frac{1}{2\Pi}\frac{M^2}{K^2} \tag{10}$$

$$E_{Cluster} = l(NE_{elec} + NE_{DA} + KE_{mp}d^4_{toBS} + NE_{elec} + NE_{fs}\frac{1}{2\Pi}\frac{M^2}{K^2}) \tag{11}$$

$$K_{opt} = \frac{\sqrt{N}}{\sqrt{2\Pi}}\sqrt{\frac{E_{fs}}{E_{mp}}}\frac{M}{d^2_{toCH}} \tag{12}$$

3.3. *The architecture of the* DCOPA *Protocol*

DCOPA is a distributed hierarchical protocol for clustering-based data communication. A competition is initiated by all nodes in the network with the computation of a timer *T(i)* based on the residual energy as well as the *DistBS* for the election of *CHs*. The nodes will just need local information. $T(i) \in]0, \tau - \delta[$, considered as a time strictly less than the duration of the period dedicated to the election of the *CHs* which is (τ). (τ) is a very small time to avoid that a node becomes a *CH* once the time of the period (τ) is finished. *DCOPA* proceeds in two phases to compose a round. In the set up phase, dedicated to the designation of the *CHs* of the current round, each node decrements its *T(i)* (see Formula 13) at the beginning of each round. If *T(i)* runs out, the node declares itself *CH* and informs all its neighbors withn a specific radius *RC* which, in turn, withdraw their candidacy for the role of *CH* by acting as normal nodes while waiting for cluster membership solicitations. The steady state phase, is divided into three periods. During the first period, the normal nodes send acknowledgment control messages to the nearest *CH*. In the second period, the *CHs* broadcast a TDMA schedule for the nodes in the cluster to send their data messages. This period is reserved for the routing of data within the cluster. During the third period, the *CHs* aggregate the data and send it to the *BS* using the MAC CSMA protocol (Carrier-Sense Multiple Access Media Access Control). Table 2 describes the variables used in the specification of *T(i)*.

Table 2: Variables of *T(i)*

Variables	Description
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.
Er_i	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.

$$T(i) = \begin{cases} (\alpha E_i + \beta D_i)(\tau - \delta) & if\ i \in G \\ \tau - \delta & \text{otherwise.} \end{cases} \tag{13}$$

$$\alpha + \beta = 1 \tag{14}$$

Where

- G is the set of nodes which were not CHs during the previous $(1/P)$ rounds,
- $P = \frac{K}{Nbr_{init}}$.
- Nbr_{init} is the initial number of nodes.

- E_i given in Formula 15 and D_i given in Formula 16, are defined as follow after the normalization process.

$$E_i = (\frac{E_{Max} - Er_i}{E_{Max}}) \tag{15}$$

$$D_i = (\frac{d_{itoBS} - d_{MintoBS}}{d_{MaxtoBS} - d_{MintoBS}}) \tag{16}$$

$$0 \leq \alpha \frac{E_{Max} - Er_i}{E_{Max}} + \beta \frac{d_{itoBS} - d_{MintoBS}}{d_{MaxtoBS} - d_{MintoBS}} < 1 \tag{17}$$

Formula (17) is verified and demonstrated in (Mir & Meziane, 2023).

4. Simulation and analysis of the impact of the energy and the *DistBS* on the performance of *DCOPA*

4.1. Network lifetime parameters

Lifetime is considered an important metric in analyzing the performance of IoT networks, particularly those based on WSNs and which are subject to energy constraints. The number of live nodes, as a function of time or rounds, is one of the most frequently used definition of network lifetime. Various descriptions and a comprehensive review of WSNs lifetime metrics are provided in (Dietrich & Dressler, 2009). In our simulations, we used the definitions for the number (or percentage) of live (or dead) nodes in the network as a function of the number of rounds (see Figure 13), namely the First Node Dead (FND), the Quarter Node Dead (QND), the Half Node Dead (HND), the Seventy-Five Percent Node Dead (SFND) and the Last Node Dead (LND).

4.2. Analysis of the energy management and the mortality rates

The network consists of (N) nodes deployed randomly in a uniform mode over an area of M*M m^2, the initial energy is the same for all nodes. Two types of messages are used, control message and data message. Table 3 illustrates the network and nodes parameters used in the simulations carried out using MATLAB [1]. The authors in *DCOPA* used ($\alpha = 0.5, \beta = 0.5$), where ($\alpha$) is the

[1] https://www.mathworks.com/

degree of importance (weight) of the energy parameter and (β) the degree of importance of the *DistBS* parameter.

Table 3: Simulation parameters

Parameters	Values	Description
$M * M$	100^2 m^2	Area network
E_{Max}	0.5 J	Initial energy
$d_{MintoBS}$	75 m	Nearest point to BS
$d_{MaxtoBS}$	183 m	Furthest point to BS
$Sink_x$	50 m	sink x-axis
$Sink_y$	175 m	sink y-axis
$MsgCtrl$	25 bytes	Control Message length
$DataMsg$	200 bytes	Data Message length
K_{opt}	5 clusters	Optimum clusters number (Heinzelman et al., 2002)

In the current simulations, the same assumptions as those made in *DCOPA* are applied. The *BS* has an unlimited energy capacity, the nodes are not dynamic and have no hardware system to know their positions, the batteries of the nodes cannot be recharged or replaced, energy exhaustion is the only cause that makes a node fail. The nodes can set their communication ranges not exceeding the maximum range. In addition to this setting, the weight of the energy and the *DistBS* will be varied. The objective of this, is to determine how energy and *DistBS* influence the selection of the *CHs* in each round through a competition based on the two parameters. Examples of combinations of (α), representing the weight of energy, and (β), representing the weight of *DistBS*, have been provided, namely the two monocriteria cases((0,1) and (1, 0)) to illustrate the relevance of the multi-criteria approach and the equitable sharing cases $(0.5, 0.5)$ and non-equitable ($(0.3, 0.7)$ and $(0.7, 0.3)$) weights. The simulation results of these combinations are illustrated in Figures 1 and 2 which contain 6 and 4 sub-figures respectively, each for a particular evaluation.

In what follows, the focus will be on the aspect of managing the overall network energy network and the nodes mortality to explain the strength of the multi-criteria approach that integrates the nodes energy as well as their *DistBS* in the clustering and data communication process.

- Figure (1a), enlarged in Figure (1b) for more clarity, illustrates the amount of the total energy in the network as a function of rounds. In the case ($\alpha = 0$, $\beta = 1$), the network conserves energy for longer periods of time and consequently obtains a better score for the Last Node Dead (LND) metric when compared to other combinations. The loss of the first nodes occurred very early. This behaviour is due, in the first place, to the election of the nodes closest to the *BS* as *CHs* because the *DistBS* is the only criterion for choosing *CHs*. The nodes closest to the *BS* win the competition for the role of *CHs* several times through a very short round spacing. This makes these nodes hot spots with high communication activity in the network. As a result, their failures would be very early, causing the first nodes to dissipate all their energy. The nodes in the middle areas that were not elected as *CHs* ensure the preservation of energy. This makes the network resistant to the last sensor metric that will dissipate all its energy. The loss of the first nodes occurred very early. This behaviour is due, firstly, to the election of the nodes closest to the *BS* as *CHs* because the *DistBS* is the only criterion for choosing *CHs*. The nodes closest to the *BS* win the competition for the role of *CHs* several times through a very short round spacing. This makes these nodes hot spots with high communication activity in the network. As a result, their failures would be very early, causing the first nodes to dissipate all their energy. The nodes in the middle areas that were not elected as *CHs* ensure the preservation of energy. This makes the network resistant to the last sensor metric that will dissipate all its energy. Secondly, the first elected *CHs* are the closest to the *BS*. By applying the concept of the minimum distance between the *CHs*, other *CHs* will be located at the end of the network, i.e. the nodes furthest away from the *BS*, this combination means that there will not be a large number of *CHs* because the distance separating the

CHs is *RC* = 50m. Clusters will be formed with a significant number of normal nodes. Consequently the number of communications with the *BS* will be reduced. For the monocriteria combination ($\alpha = 1$, $\beta = 0$), the network shows a performance in the FND metric. Once the loss of the first nodes appears, the mortality would be considerable, i.e. in few rounds the network undergoes a significant failure of its membership. This accelerates the loss of the whole network, resulting in the lowest LND of any other combination. The consequences described above are caused by the fact that the nodes with more energy assume the role of *CHs* because all the nodes have enough energy (the first few rounds) and that the *DistBS* does not intervene in the election condition of *CHs*. The nodes are more resistant to failure. A minor energy difference may be a factor in having *CHs* nodes that will be different from one round to another. These elements can retard the achievement of FND. Nodes which are far from the *BS* can proclaim themselves as *CHs*. Ensuring the aggregation and sending of data to the *BS* over very long distances as well as the management of cluster join messages, cause their failures which engenders an LND very early. These performances are due to the fact that this combination is monocriterion where there is just the consideration of energy as the only decisive factor for a node to proclaim itself *CH*. In cases where both criteria share weights, the results are balanced between better energy management and better mortality rate compared to the previous two monocriteria cases. This supports the multi-criteria consideration of the authors of *DCOPA* and more specifically the equitable sharing ($\alpha = 0.5$, $\beta = 0.5$) of the influence of the energy and *DistBS* criteria due to their very close relationship.

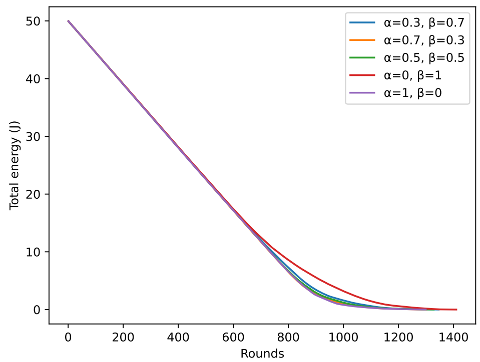
- Figure (1c), zooming into Figure (1d) for better visibility of the results, shows the average energy per node as a function of the rounds. The results show that the 5 combinations behave similarly up to round 600. However, it should be noted that the combination ($\alpha = 0$, $\beta = 1$) resulted in the highest average residual energy per node. The combination ($\alpha = 1, \beta = 0$) has the worst result of all combinations. The non consideration of the parameter *DistBS* has a very negative influence on the results, this can be justified by the fact that choosing nodes as *CHs* and which are very far from the *DistBS* generates an excessive consumption of energy of the *CHs* and what is directly reflected on the average energy of the totality of the nodes of the network. The other combinations, including ($\alpha = \beta = 0.5$), obtained results between the two monocriteria cases ($\alpha = 0$,$\beta = 1$) and ($\alpha = 1$,$\beta = 0$).

- Figure (1e), zooming into Figure (1f) for clarity, gives us information on the average energy consumed by a node as a function of rounds. When examining the results, it becomes apparent that if the *DistBS* is discarded from the competition function for the role of *CH*, there will be a noticeable improvement in the average consumption compared to the other cases of combinations with stability and balancing in terms of average energy per node. This also shows a balanced consumption throughout the lifetime of the network. In the case of considering the *DistBS* as the only criterion, this results in an average energy per node that is not stable with a variation that is considered unbalanced. This is due to the fact that there are nodes which are closer to the *BS* and which are highly solicited in the role of the *CH* independently of their residual energies. This constitutes hot spots in the network and an unbalanced load distribution. In the cases where (α) and (β) are varied, there will be average variations between the two monocriteria cases.

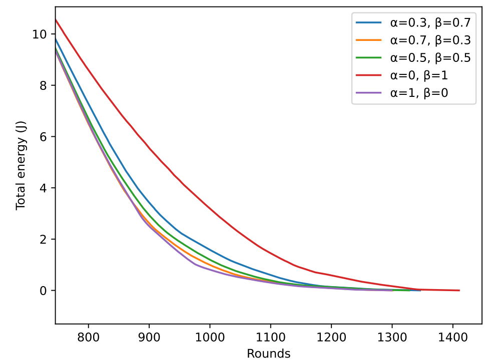
- Figure (2a), zooming into Figure (2b) for more precision and visibility of the results, shows the average energy consumed by the network as a function of the rounds. It is noticeable that in the case of not considering the energy in the election equation of *CHs*, the average consumption of the network varies in an interval of values which is wider than the other cases until round 400. A slight decrease is observed in the interval between rounds 400 and 600. Starting from round 600, the consumption decreases in a considerable and accelerated way. This allows us to conclude that the absence of the consideration of energy as one of the factors in the election of *CH* generates an unstable and unbalanced consumption across the rounds. This is due to the hot spots which are the nodes bordering the network which and closest to the *BS*, i.e. they are highly solicited for the role of *CH*, this excludes the nodes in the middle for the role of *CHs* because of the considered clustering radius in *DCOPA*. The case where considering just the energy as the main factor in the equation, the balancing is maintained until round 800 with a slight variation. From the round 800, a gradual decrease is observed, which is not abrupt compared to the previous case. For the cases of variation of (α) and (β), the average consumption is between the two preceding cases.

- Figure (2c), zooming into Figure (2d) in order to visualise better the variations of the graphs, gives us an indication of the mortality of the nodes in the network. For ($\alpha = 0$, $\beta = 1$), i.e. the energy is excluded in the selection process of *CHs*, the mortality and the loss of nodes is very considerable and sudden compared to other combinations. This sudden loss of nodes continues until round 900 where there is an improvement in mortality compared to other cases. In the case of the combination ($\alpha = 1$, $\beta = 0$), the obtained results are very satisfactory compared to other cases until around round 900 where the network loses its performance in terms of mortality rate comparing it to all the other combinations. In the cases where (α) and (β) share the weights, the performance of the network is between the two extreme cases (monocriteria). This justifies the multi-criteria approach to preserve the two positive impacts of including energy and *DistBS* in the competition of election of *CHs* and the formation of their clusters.

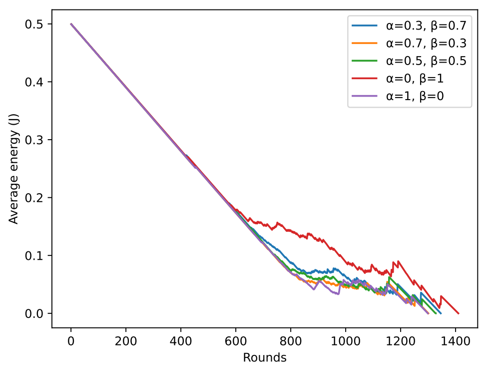
Discussion 1: Through the simulations, some very interesting conclusions were extracted, in particular the degree of influence of the weights of the criteria



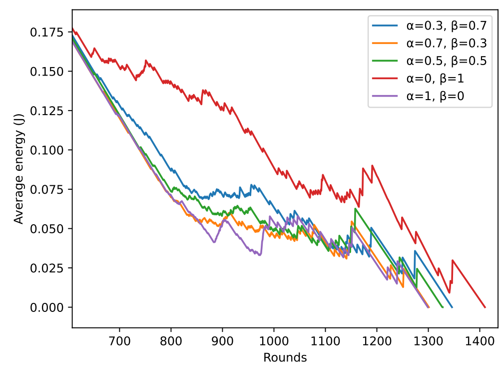
(a) Total network energy as a function of rounds



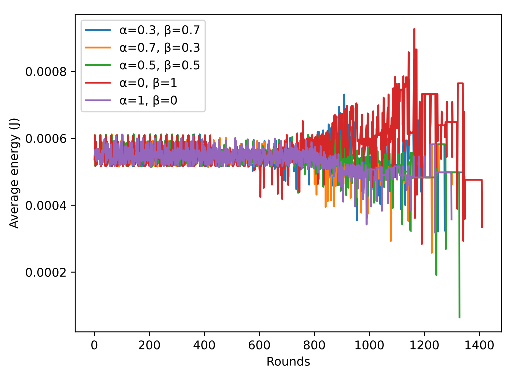
(b) Total network energy as a function of rounds



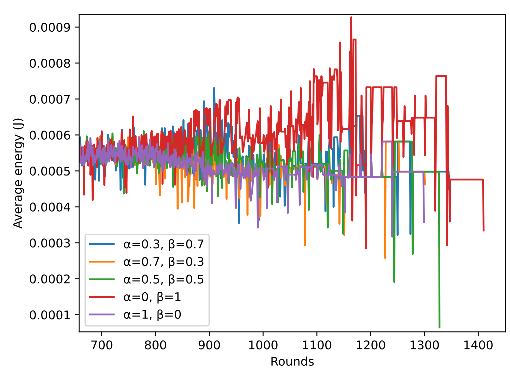
(c) Average residual energy per node as a function of rounds



(d) Average residual energy per node as a function of rounds

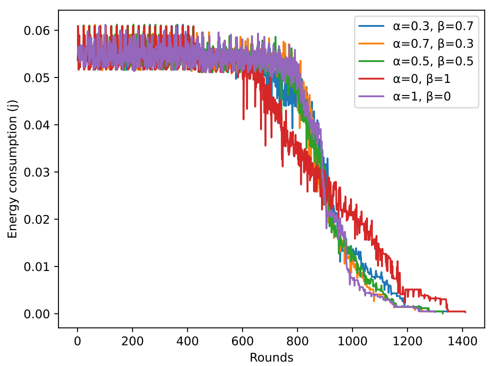


(e) Average energy consumed by a node as a function of rounds

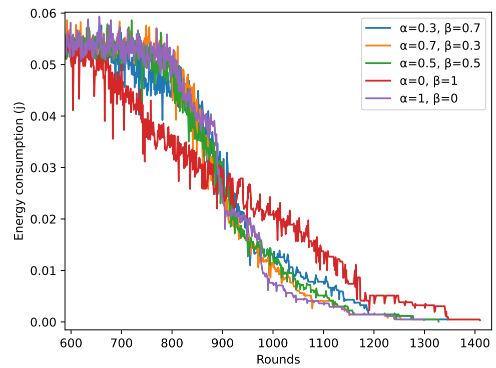


(f) Average energy consumed by a node as a function of rounds

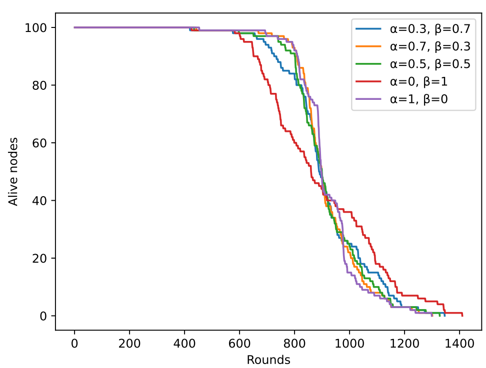
Figure 1: *DCOPA* performances With (α, β) variation
(Total network energy and Average node's energy (consumed and residual))



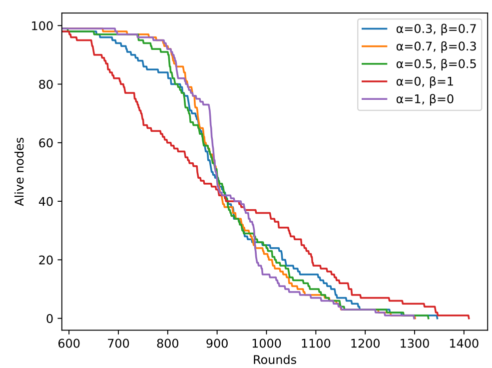
(a) Average network energy consumption as a function of rounds



(b) Average network energy consumption as a function of rounds



(c) Number of alive nodes as a function of rounds



(d) Number of alive nodes as a function of rounds

Figure 2: *DCOPA* performances With (α, β) variation.
Average network energy consumption and Number of alive nodes

(energy and *DistBS*) in the election competition of the set of *CHs* in the network and their impacts on the performance of the network in terms of energy management of the nodes and the network, load balancing, rotation of the role of the *CH* between nodes and the degree of loss of nodes in the network. The aspect of global network energy management and node mortality is emphasized in our conclusions to explain the advantage of the multi-criteria approach that integrates the energy of the nodes as well as their *DistBS* into the clustering process. It was noted that the network preserves the energy with the combination ($\alpha = 0$, $\beta = 1$), this is the result of the election of the nodes close to the *BS* as *CHs*. With the application of the concept of the minimal distance between the *CHs* there will be other *CHs* at the extremity of the network. This ensures that there is not a large number of *CHs* because of the constraint of the minimum distance between two *CHs* (greater than 50 m in our case). There will be clusters with a large number of normal nodes. This means that the number of communications with the *BS* is reduced. As the same nodes can assume the role of *CH* several consecutive times, early nodes will be present that deplete their energies very early and will have a high mortality. On the other hand, nodes that did not obtain the role of *CH* that are in the middle regions of the network can preserve their energies and improve the LND metric accordingly. Concerning the combination ($\alpha = 1, \beta = 0$), a better energy balancing result will be obtained. This keeps the network in its entirety for more time than the previous cases. After this stable period, the network is not impacted abruptly by the mortality rate. In the cases where the two criteria share the weights, the results are balanced between a good energy management and a better mortality, which supports the consideration and the choice of the authors of *DCOPA* for the combination ($\alpha = 0.5, \beta = 0.5$). This is clearly illustrated in Figures 1 and 2.

4.3. Simulations and analysis of scalability performance

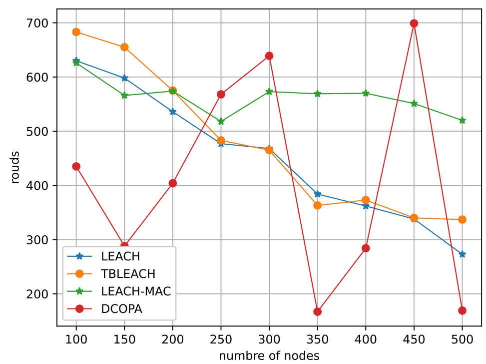
Scalability is one of the most important properties of any protocol, especially those dedicated to clustering and routing in *IoT* networks because of the large number of devices and nodes in their applications. This means that if a protocol does not guarantee scalability, it has a significant drawback. In what follows, this aspect is evaluated for the *DCOPA* protocol with different weights of the criteria involved in the selection of *CHs*, namely energy and *DistBS*. A total of 9 networks were generated, ranging from 100 to 500 nodes with an increment of 50 nodes for each network, deployed under the same previously stated conditions. Three combinations of (α) and (β) were simulated for the 9 networks. The results were then compared to the same protocols used in (Mir & Meziane, 2023), namely LEACH, LEACH-MAC and TB-LEACH in terms of network lifetime for which The FND, QND, HND, SFND, and LND were chosen. The simulations results shown in Figures 3, 4 and 5 for the combinations ($\alpha = 0.5, \beta = 0.5$), ($\alpha = 0.3, \beta = 0.7$) and ($\alpha = 0.7, \beta = 0.3$) respectively, illustrate that the *DCOPA* protocol supports scalability compared to other protocols.

Discussion 2: As observed through all the sub-figures of all the Figures (3, 4 and 5) presenting FND, QND, HND, SFND and LND for the different sizes of the network (initial number of nodes), *DCOPA* shows that it outperforms other protocols in terms of scalability by keeping its energy conservation properties and better lifetime in all the metrics considered. Concerning LEACH and TB-LEACH, the results show that they do not support scalability. As the number of nodes increases, these protocols exhibit a decrease in lifetime metrics. LEACH-MAC shows through the results its support of scalability by keeping this performance for the considered metrics. In addition to the fact that energy and *DistBS* are involved in the competition of nodes for the role of the *CH*, there are other well-founded and convincing arguments that *DCOPA* is a scalable protocol and that this performance will not be affected in large-scale IoT networks. The arguments presented below apply to both combinations with equitable ($\alpha = \beta = 0.5$) and non-equitable (($\alpha = 0.3, \beta = 0.7$) and ($\alpha = 0.7, \beta = 0.3$)) criteria weights. The main properties of *DCOPA* ensure that increasing the number of nodes also causes an increase in the number of competing nodes. To avoid having a small number of clusters with a large number of normal nodes, the optimal number of *CHs* must also increase. Similarly, to avoid the formation of large clusters and a large inter-cluster communication radius, the clustering radius must decrease as the number of nodes increases. *DCOPA* is less efficient in terms of the FND metric, this is the result of the re-election of some isolated nodes as *CHs* for many rounds consecutively until their total exhaustion. These nodes were not solicited by other *CHs* nodes although their competition times are pushed towards $(\tau - \delta)$ which makes their energy consumption very considerable in each round.

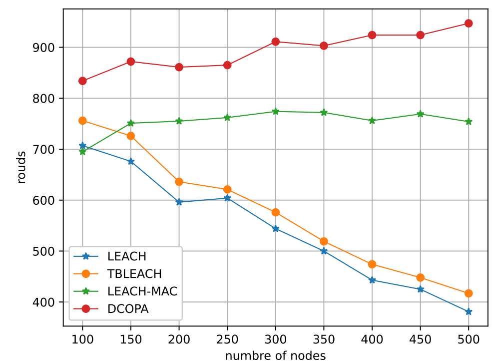
4.4. *Analysis of clusters load balancing and distribution of* CHs *and their member nodes*

A clustering protocol in IoT environments need to take into consideration other very important metrics such as load balancing between clusters in terms of the number of nodes, the distribution of member nodes and the area of its extent as well as its shape and the distribution of *CHs* along the entire deployment area. If these properties or metrics are ensured by a clustering protocol, inevitably, there will be a better management of the energy constraints related to the activities of the clustering process and the data communication processes. This means that a better energy efficiency and a prolonged lifetime of the network are obtained. In this same context, the Figures 6, 7, and 8 are used to illustrate the load balancing of the clusters, their density, their distribution as well as the positioning of the *CHs* across the network. The exploration of these properties was performed for 100 nodes in Figure 6, 300 nodes in Figure 7, and 500 nodes in Figure 8. The parameters $(\alpha = 0.5, \beta = 0.5)$ were maintained. The objective of repeating the same experiments and simulations, varying the number of nodes, is the checking of the scalability support at the clusters load balancing level, the clusters density, the number of clusters and the distribution balancing of the *CHs* at the scaling up. This property is a key element of IoT applications, in particular those based on wireless sensor networks. The following will provide a description of the Figures 6, 7, and 8. Subsequently, the conclusions will be synthesized and the factors that impact their results will be discussed.

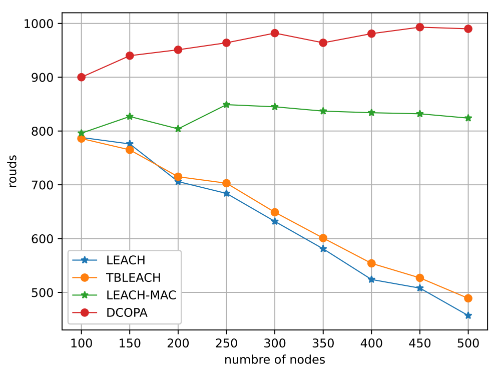
- It is clearly apparent from Figures ($6a$), ($7a$), and ($8a$) for the 100, 300, and 500 nodes, respectively, that it is possible to distinguish between the different clusters by their colours and shape as well as by the distribution



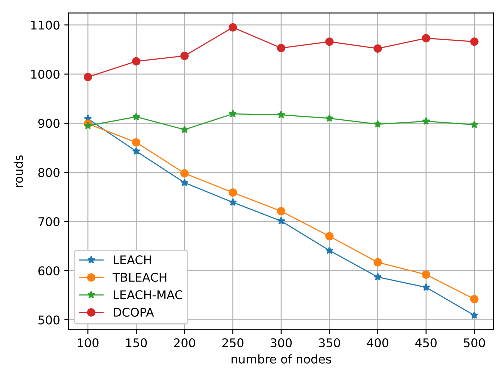
(a) First Node Dead (FND)



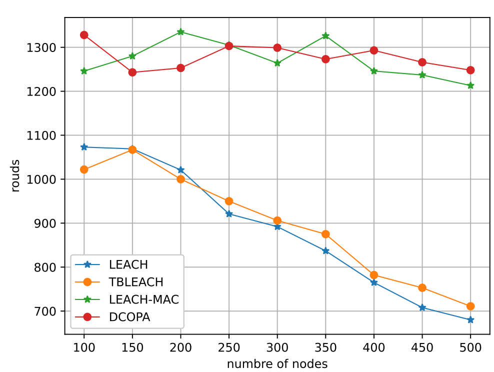
(b) Quarter of the Nodes Dead (QND)



(c) Half Nodes Dead (HND)

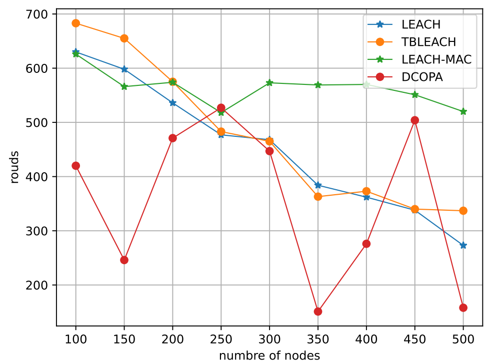


(d) Seventy Five percent Nodes Dead (SFND)

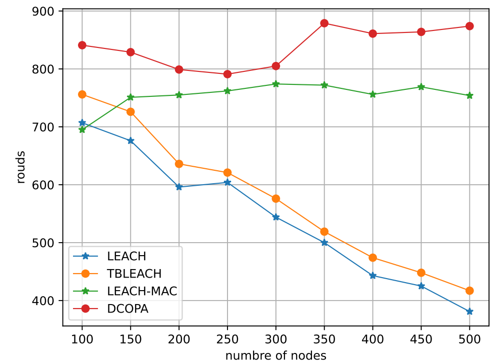


(e) Last Node Dead (LND)

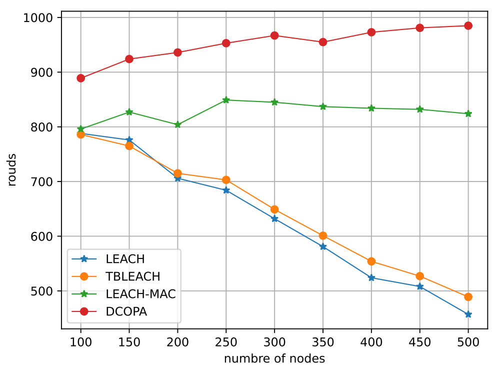
Figure 3: Percentage Nodes Dead as a function of initial number of nodes for ($\alpha = 0.5$) and ($\beta = 0.5$)



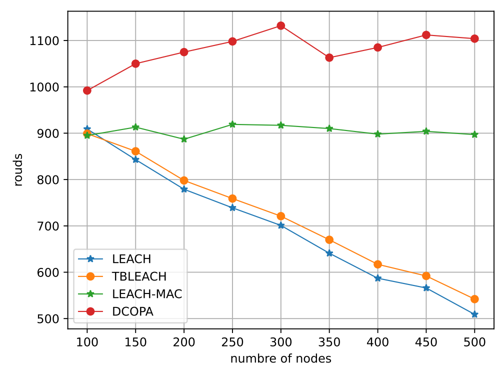
(a) First Node Dead (FND)



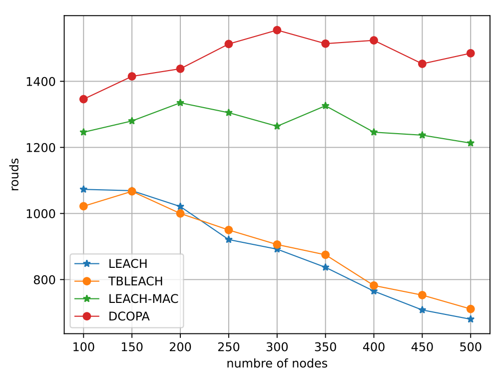
(b) Quarter of the Nodes Dead (QND)



(c) Half Nodes Dead (HND)

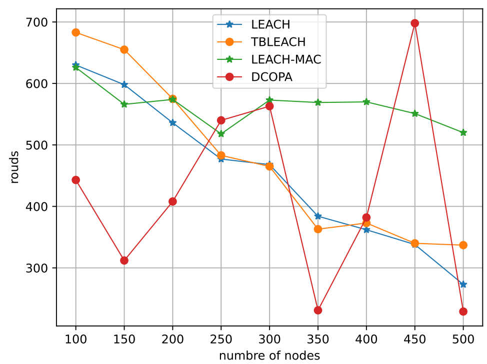


(d) Seventy Five percent Nodes Dead (SFND)

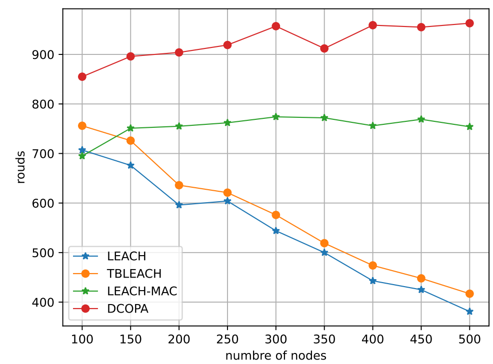


(e) Last Node Dead (LND)

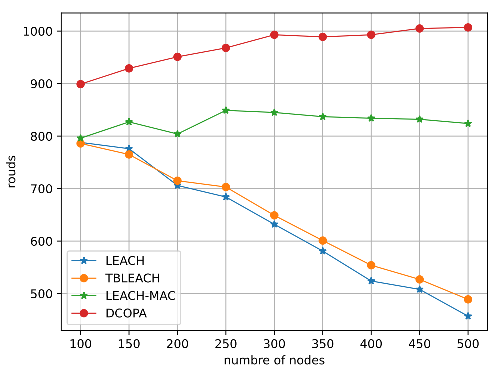
Figure 4: Percentage Nodes Dead as a function of initial number of nodes for ($\alpha = 0.3$) and ($\beta = 0.7$)



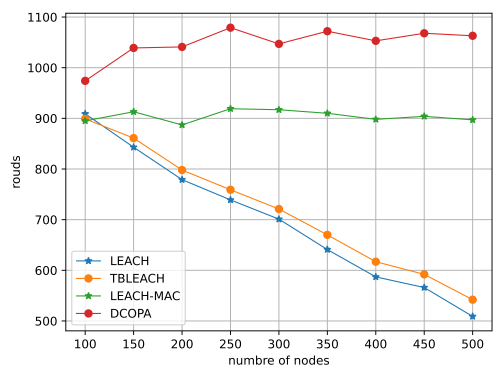
(a) First Node Dead (FND)



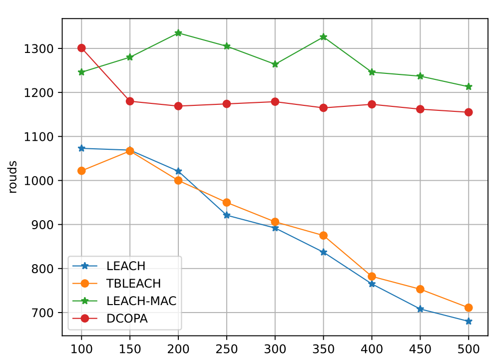
(b) Quarter of the Nodes Dead (QND)



(c) Half Nodes Dead (HND)



(d) Seventy Five percent Nodes Dead (SFND)



(e) Last Node Dead (LND)

Figure 5: Percentage Nodes Dead as a function of initial number of nodes for ($\alpha = 0.7$) and ($\beta = 0.3$)

of the *CHs* in the deployment surface. In addition, the degree of concentration or density of nodes within the same cluster in well-defined areas can also be observed.

- In Figures (6*b*), (7*b*) and (8*b*) for respectively 100, 300 and 500 nodes, linear plots, which are the result of applying the concept of multiple linear regression(Saporta, 2006), have been carried out. It consists in applying the linear regression for each cloud of points of the same colour or of the same cluster in our context. This regression aims to graphically represent the most appropriate straight line to be applied to a cloud of points from the same cluster by minimising the sum of the squares of the distances between the points and the line. This makes it possible to check for the presence or absence of nodes that are far from the line drawn.

- The pairs of Figures (6*c*, 6*d*), (7*c*, 7*d*) and (8*c*, 8*d*) for 100, 300 and 500 nodes respectively, plot the joint distribution of two variables X and Y, which represent the position of each node in the network, adding marginal axes that show the univariate distribution of each separately.

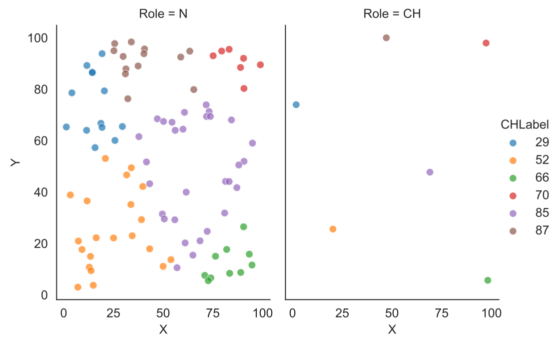
- Figures (6*e*), (7*e*) and (8*e*) for 100, 300 and 500 nodes, respectively, show the residual energies of the nodes that did not dissipate a large amount of energy because they do not have any communication to the BS, contrary to the *CHs* that communicate their aggregated data.

- Figures (6*f*), (7*f*) and (8*f*) for respectively 100, 300 and 500 nodes, show the *DistBS* of the nodes and the *CHs* that form a same cluster.

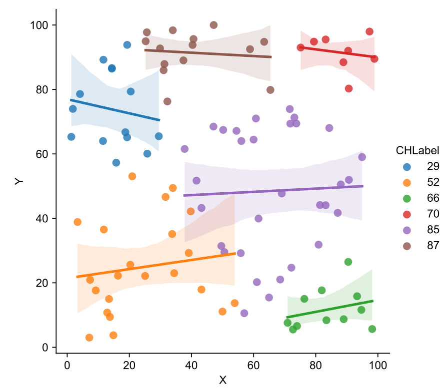
Discussion 3: The very interesting properties of *DCOPA* are highlighted from simulations of its functioning and architecture. These properties concern key aspects of any clustering protocol. The results shown in Figures 6, 7 and 8 illustrate that the *DCOPA* protocol ensures a homogeneous distribution of *CHs* and clusters on the deployment surface and a load balancing of clusters on the density aspect in terms of number of member nodes (distribution). These results are well maintained even when increasing the number of nodes in the network, which attributes to *DCOPA* the property of scalability. The achievement of these characteristics is due to the functioning of *DCOPA* and its architecture. All the nodes of the network participate in a competition to win the role of *CH*. Nodes with a minimum weighted sum have more chance to declare themselves as *CH*, then broadcast an announcement message on a Communication Radius called *RC*, see (Mir & Meziane, 2023), which is calculated based on the optimum number of *CHs* in the network which is sensitive of the initial number of nodes. The value of *RC* decreases as the number of nodes increases to obtain a greater number of clusters, to avoid clusters which contain a sizeable number of nodes and to distribute the load of member nodes over a large number of clusters. Figures (6*c*, 6*d*), (7*c*, 7*d*) and (8*c*, 8*d*) for 100, 300 and 500 nodes respectively, show this highly connected relationship between the number of initial nodes, the *RC*, the number of clusters and the load distribution of member nodes and *CHs*. The support of scalability is well demonstrated by *DCOPA* by providing a smooth distribution of *CHs* and clusters. In conclusion, *DCOPA* produces a virtual and dynamic grid of *CHs* and their clusters by analysing the Figures (6*a*), (7*c*), (8*a*) for respectively 100, 300 and 500 nodes.

4.5. Covariance and linear correlation

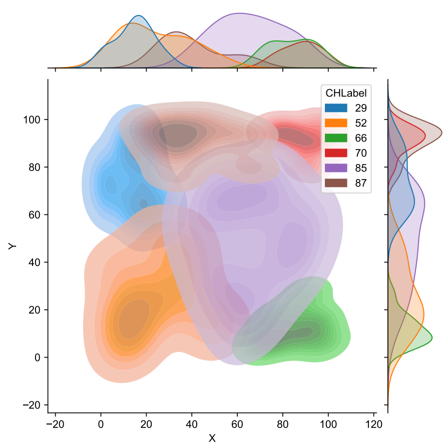
In what follows, it is necessary to define the statistical elements required for the analysis of the impact of the initial number of nodes and the variation of (α, β) on the defined lifetime parameters. To better understand the influence of one parameter on another, the linear correlation coefficient and the covariance of two statistical series (Saporta, 2006) were used.



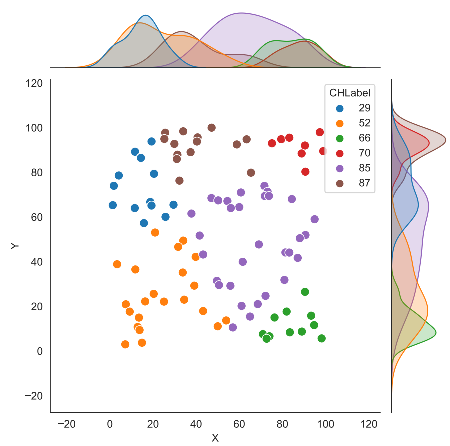
(a) Distribution of *CHs* and normal nodes in the deployment surface



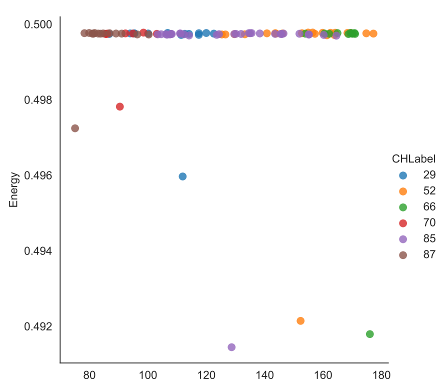
(b) Cluster analysis using multiple linear regression



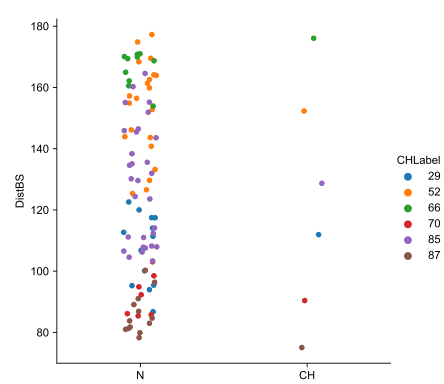
(c) Joint distribution and univariate distribution (Clusters)



(d) Joint distribution and univariate distribution (Nodes)

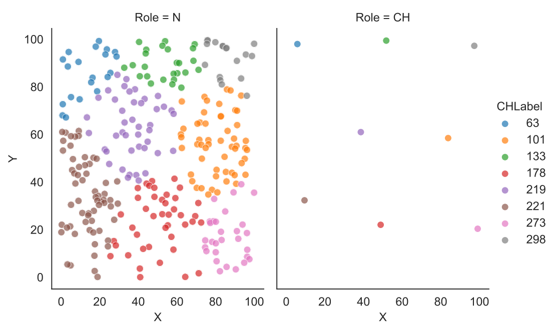


(e) The residual energies of the *CHs* nodes and the normal nodes that form the same cluster

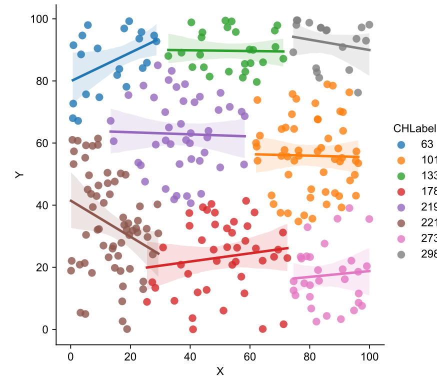


(f) The *DistBS* of *CHs* nodes and normal nodes that form the same cluster

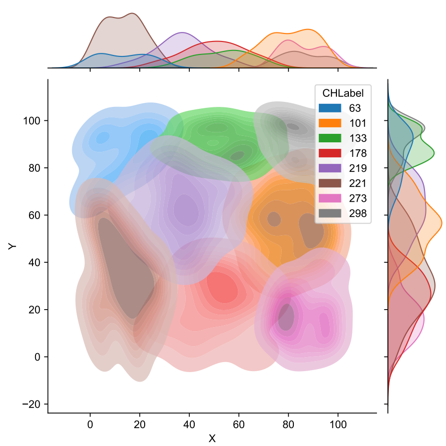
Figure 6: Analysis of clusters load balancing and distribution of *CHs* and their member nodes (100 nodes)



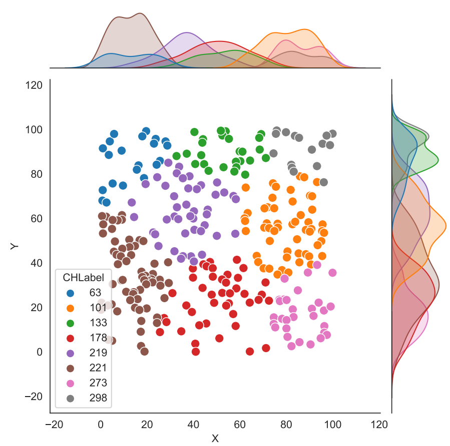
(a) Distribution of *CHs* and normal nodes in the deployment surface



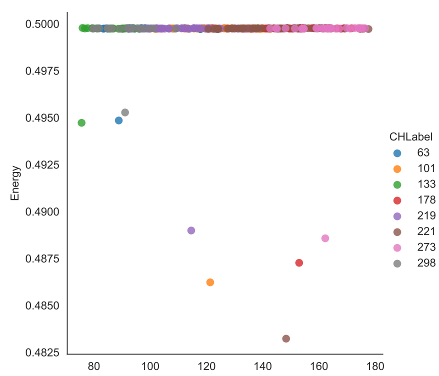
(b) Cluster analysis using multiple linear regression



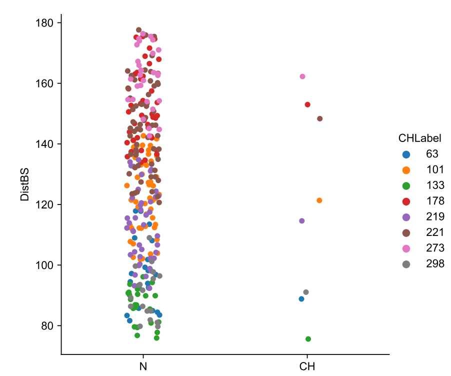
(c) Joint distribution and univariate distribution (Clusters)



(d) Joint distribution and univariate distribution (Nodes)

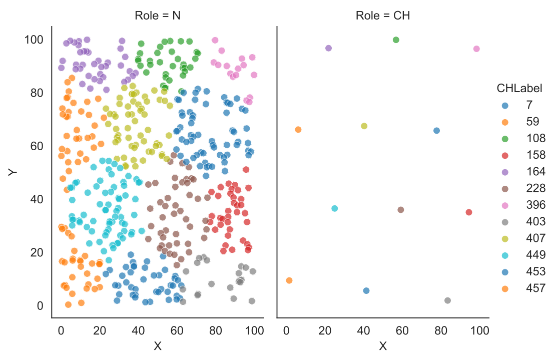


(e) The residual energies of the *CHs* nodes and the normal nodes that form the same cluster

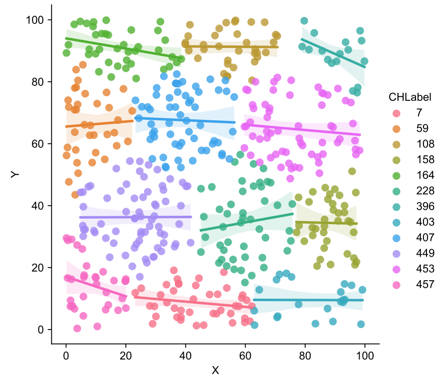


(f) The *DistBS* of *CHs* nodes and normal nodes that form the same cluster

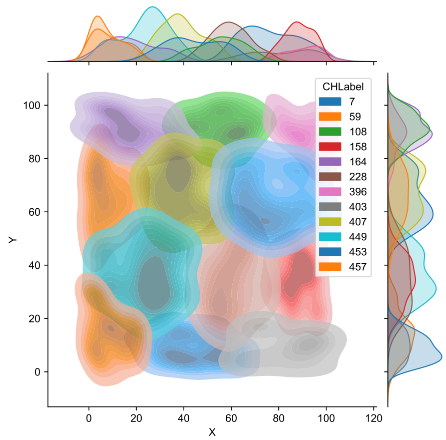
Figure 7: Analysis of clusters load balancing and distribution of *CHs* and their member nodes (300 nodes).



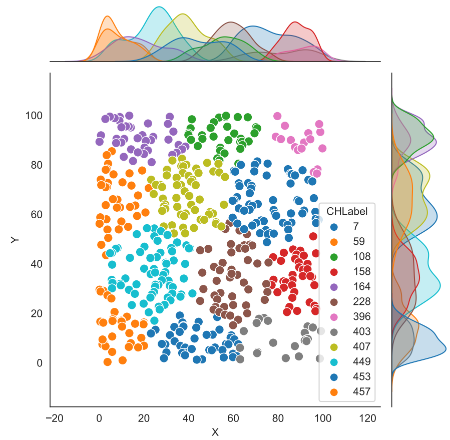
(a) Distribution of *CHs* and normal nodes in the deployment surface



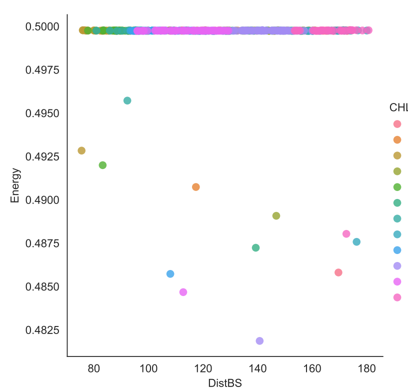
(b) Cluster analysis using multiple linear regression



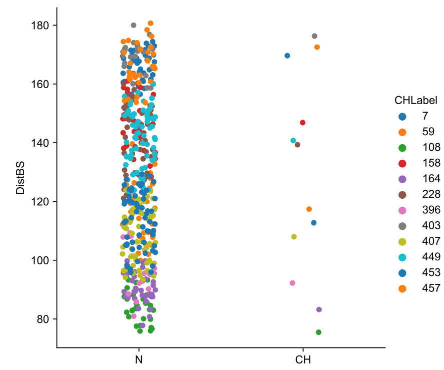
(c) Joint distribution and univariate distribution (Clusters)



(d) Joint distribution and univariate distribution (Nodes)



(e) The residual energies of the *CHs* nodes and the normal nodes that form the same cluster



(f) The *DistBS* of *CHs* nodes and normal nodes that form the same cluster

Figure 8: Analysis of clusters load balancing and distribution of *CHs* and their member nodes (500 nodes)

4.5.1. Covariance of two statistical series

The covariance of two statistical series $x = x_1, x_2, \ldots, x_n$ and $y = y_1, y_2, \ldots, y_n$ having the same number of elements is defined by:

$$Cov(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \tag{18}$$

$$= \frac{1}{n-1}[\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}] \tag{19}$$

$$= \frac{1}{n-1}[\sum_{i=1}^{n} x_i y_i - n\overline{xy}] \tag{20}$$

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + ... + x_n}{n} \tag{21}$$

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{y_1 + y_2 + ... + y_n}{n} \tag{22}$$

This is the average of the products of the deviations of the values from the mean of each series.

- If x and y are positively associated, then $Cov(x, y)$ will be large and positive
- If x and y are negatively associated, then $Cov(x, y)$ will be large and negative
- If the variables are not positively nor negatively associated, then $Cov(x, y)$ will be small

4.5.2. Linear correlation coefficient

The linear correlation coefficient of two statistical series is defined by : r

$$r = \frac{1}{n-1}\sum_{i=1}^{n}(\frac{x_i - \overline{x}}{s_x})(\frac{y_i - \overline{y}}{s_y}) \tag{23}$$

$$s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}} \quad and \quad s_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n-1}} \tag{24}$$

$$r = \frac{Cov(x,y)}{s_x s_y} \tag{25}$$

The linear corelation coefficient quantifies the strength of the linear relationship between x and y

- Always falls between -1 and $+1$
- A positive r value indicates a positive association
- A negative r value indicates a negative association
- r value close to $+1$ or -1 indicates a strong linear association
- r value close to 0 indicates a weak association.

4.5.3. Correlation matrix

When there are several variables defined on the same set of individuals, often, the calculation of all correlation coefficients between variables taken two by two is desired. It is useful to gather these coefficients in a single table (matrix) of which :

- The rows and columns represent the variables.
- The intersection of the row l and the column k, is r(l,k), correlation coefficient between the variables l and tk.

This table is symmetrical: $r(k,l) = r(l,k)$. Its diagonal has only the value 1: $r(k,k) = 1$.

	alpha	*beta*	*FND*	*QND*	*HND*	*SFND*	*LND*
alpha	1						
beta	-1	1					
FND	0.464682	-0.46468	1				
QND	0.852182	-0.85218	0.27123	1			
HND	0.556355	-0.55636	0.393084	0.726483	1		
SFND	-0.75399	0.753987	-0.28545	-0.91548	-0.83547	1	
LND	-0.86921	0.869208	-0.58465	-0.92572	-0.79682	0.863776	1

Figure 9: Correlation Matrix of lifetime parameters for a network of 100 nodes depending on the (α, β) variation

4.6. Analysis of the impact of the variation of the initial number of nodes and the weights (α, β) on the lifetime parameters

4.6.1. Setting the initial number of nodes and varying the weights (α, β)

Simulations were performed on networks of varying sizes in terms of the initial number of nodes (NbrNdInit), ranging from 100 to 500 nodes with an increment of 50 nodes at each iteration. For each network, the variations of the (α, β) combinations were examined as well as the initial number of nodes in order to evaluate the simultaneous impact of these two parameters on the lifetime parameters such as FND, QND, HND, SFND and LND. In other words, the lifetime parameters were studied as a function of scalability and the interest given to each criterion included in the selection process of the *CHs*. The lifetime parameters obtained for networks of 100, 300, and 500 nodes with variations of the (α, β) combinations are given in Tables 4, 5, and 6, respectively. Next, the correlation matrices of the lifetime parameters as a function of the variation of (α, β) were presented. These matrices are shown in Figures 9, 10, and 11 for networks of 100, 300, and 500 nodes, respectively. Mathematical and statistical methods mentioned and explained in sections 4.5.1, 4.5.2, and 4.5.3 were used to analyze the data and determine the correlation between the different parameters considered.

Discussion 4: As so well illustrated by the values given in the Tables 4, 5 and 6 and by the correlation coefficients defined in Figures 9, 10 and 11, it is observed, firstly, that scalability is well ensured. The number of nodes in the network does not influence the results of the correlation coefficients in all combinations

Table 4: Lifetime parameters obtained based on the variation of (α, β) weights with 100 nodes

		Lifetime Parameters				
NbrNdInit	(α, β) VALUES	FND	QND	HND	SFND	LND
100	$(0, 1)$	426	733	860	1071	1410
	$(0.1, 0.9)$	430	749	890	1024	1390
	$(0.2, 0.8)$	428	806	889	983	1362
	$(0.3, 0.7)$	420	841	889	992	1346
	$(0.4, 0.6)$	423	834	899	990	1332
	$(0.5, 0.5)$	435	834	900	994	1328
	$(0.6, 0.4)$	438	848	897	985	1323
	$(0.7, 0.3)$	443	855	899	974	1301
	$(0.8, 0.2)$	454	851	895	976	1304
	$(0.9, 0.1)$	412	860	887	976	1338
	$(1, 0)$	454	858	897	977	1299

Table 5: Lifetime parameters obtained based on the variation of (α, β) weights with 300 nodes

		Lifetime Parameters				
NbrNdInit	(α, β) VALUES	FND	QND	HND	SFND	LND
300	$(0, 1)$	283	645	904	1301	1717
	$(0.1, 0.9)$	349	704	912	1233	1738
	$(0.2, 0.8)$	432	751	931	1199	1694
	$(0.3, 0.7)$	447	805	967	1132	1555
	$(0.4, 0.6)$	605	869	974	1083	1393
	$(0.5, 0.5)$	639	911	982	1053	1299
	$(0.6, 0.4)$	525	932	991	1044	1230
	$(0.7, 0.3)$	563	957	993	1047	1179
	$(0.8, 0.2)$	599	964	994	1052	1141
	$(0.9, 0.1)$	632	968	1000	1048	1111
	$(1, 0)$	632	971	1007	1049	1090

	alpha	*beta*	*FND*	*QND*	*HND*	*SFND*	*LND*
alpha	1						
beta	-1	1					
FND	0.860312	-0.86031	1				
QND	0.949265	-0.94927	0.933077	1			
HND	0.933457	-0.93346	0.922107	0.983536	1		
SFND	-0.87618	0.87618	-0.9411	-0.97929	-0.97686	1	
LND	-0.96874	0.968736	-0.88797	-0.97534	-0.95822	0.923369	1

Figure 10: Correlation Matrix of lifetime parameters for a network of 300 nodes depending on the (α, β) variation

Table 6: Lifetime parameters obtained based on the variation of (α, β) weights with 500 nodes

		Lifetime Parameters				
NbrNdInit	(α, β) VALUES	FND	QND	HND	SFND	LND
500	$(0, 1)$	159	646	998	1317	1863
	$(0.1, 0.9)$	150	746	978	1213	1853
	$(0.2, 0.8)$	152	806	972	1156	1700
	$(0.3, 0.7)$	158	874	985	1104	1485
	$(0.4, 0.6)$	181	914	994	1074	1346
	$(0.5, 0.5)$	169	947	990	1066	1248
	$(0.6, 0.4)$	218	963	995	1064	1200
	$(0.7, 0.3)$	229	963	1007	1063	1155
	$(0.8, 0.2)$	208	968	1004	1063	1129
	$(0.9, 0.1)$	229	973	1013	1061	1102
	$(1, 0)$	227	974	1016	1063	1093

	alpha	*beta*	*FND*	*QND*	*HND*	*SFND*	*LND*
alpha	1						
beta	-1	1					
FND	0.860312	-0.86031	1				
QND	0.949265	-0.94927	0.933077	1			
HND	0.933457	-0.93346	0.922107	0.983536	1		
SFND	-0.87618	0.87618	-0.9411	-0.97929	-0.97686	1	
LND	-0.96874	0.968736	-0.88797	-0.97534	-0.95822	0.923369	1

Figure 11: Correlation Matrix of lifetime parameters for a network of 500 nodes depending on the (α, β) variation

of (α, β). This reveals a very interesting property of *DCOPA*, which is the maintenance of performance in the case of scaling up. Secondly, for the three networks, the (α) weight of energy has a strong positive linear association with the parameters FND and HND and a strong negative linear association with the parameters SFND and LND. This means that when the (α) weight of energy increases, the values of FND and HND also increase while the values of SFND and LND decrease. This relationship is observed through the linear correlation coefficients and correlation matrices obtained. For the weight (β) associated with the criterion of the *DistBS*, it is exactly the opposite when comparing it with the correlation of (α) with all lifetime parameters. Secondly, for the three networks, the (α) weight of energy has a strong positive linear association with the parameters FND and HND and a strong negative linear association with the parameters SFND and LND. This means that when the (α) weight of energy increases, the values of FND and HND also increase while the values of SFND and LND decrease. This relationship is observed through the linear correlation coefficients and correlation matrices obtained. For the weight (β) associated with the criterion of the *DistBS*, it is exactly the opposite when comparing it with the correlation of (α) with all lifetime parameters. this can be explained by the fact that when more interest is assigned to energy it positively affects the lifetime parameters, namely FND, QND and HND. This is due to the fact that the nodes have enough energy at the launch of the network. The high probability of electing nodes far away from the *BS* as *CHs* does not lead to their failures, i.e., they have enough energy to withstand data communications over large distances. The fact that the network undergoes a considerable loss of energy and nodes after a significant number of rounds causes a situation where the subsequently selected nodes may be at a considerable distance from the *BS*, which accelerates their failure. For the variation of (β), which is associated with the criterion of the *DistBS*, the correlation results are exactly the opposite of those of (α). This is due to the fact that favouring the criterion of *DistBS* over the energy criterion generates hotspots which are the nodes closest to the *BS* that will frequently be elected as *CHs* which engenders the gradual failure of the network starting from the nodes of the closest regions of the BS to the nodes of the most distant regions. Excluding the more energetic nodes that are further away from the BS from the role of the CH, generates a progressive propagation of node failure in the opposite direction of the BS. This allows us to obtain a strong positive linear correlation with the SFND and LND parameters.

4.6.2. Setting the weights (α, β) and varying the initial number of nodes

The linear correlation coefficients were calculated (the Correlation Matrix) between the initial number of nodes parameter with the lifetime parameters to measure the strength and direction of their linear relationship with all (α, β) variations considered. Each network, ranging from 100 to 500 nodes with an increment of 50 nodes at each iteration, is simulated with a set of (α, β) combinations. The results of the correlation coefficients are presented in Figure 12.

alpha	beta	alpha	beta	alpha	beta	alpha	beta	alpha	beta	alpha	beta
0	1	0.1	0.9	0.2	0.8	0.3	0.7	0.4	0.6	0.5	0.5
	NbrNdInit		*NbrNdInit*		NbrNdInit		*NbrNdInit*		NbrNdInit		*NbrNdInit*
NbrNdInit	1	NbrNdInit	1	NbrNdInit	1	NbrNdInit	1	NbrNdInit	1	NbrNdInit	1
FND	-0.7284	FND	-0.5221	FND	-0.37527	FND	-0.31961	FND	-0.1903	FND	-0.10956
QND	-0.54114	QND	0.27053	QND	0.321312	QND	0.60795	QND	0.891275	QND	0.950689
HND	0.837462	HND	0.854815	HND	0.964249	HND	0.93368	HND	0.941336	HND	0.896137
SFND	0.884706	SFND	0.761829	SFND	0.632338	SFND	0.685194	SFND	0.756168	SFND	0.666174
LND	0.95278	LND	0.9389	LND	0.838279	LND	0.591557	LND	0.345075	LND	-0.31682

alpha	beta	alpha	beta	alpha	beta	alpha	beta	alpha	beta
0.6	0.4	0.7	0.3	0.8	0.2	0.9	0.1	1	0
	NbrNdInit		*NbrNdInit*		NbrNdInit		*NbrNdInit*		NbrNdInit
NbrNdInit	1	NbrNdInit	1	NbrNdInit	1	NbrNdInit	1	NbrNdInit	1
FND	0.017463	FND	-0.01712	FND	-0.09654	FND	0.016915	FND	-0.00818
QND	0.933301	QND	0.880091	QND	0.860645	QND	0.852339	QND	0.843565
HND	0.880826	HND	0.937474	HND	0.917377	HND	0.906534	HND	0.898991
SFND	0.619003	SFND	0.672747	SFND	0.685209	SFND	0.78179	SFND	0.836529
LND	-0.4512	LND	-0.65502	LND	-0.72407	LND	-0.70966	LND	-0.72804

Figure 12: Correlation matrix for the NbrNdInit with lifetime parameters as a function of the weights (α, β)

Discussion 5: The correlation matrix presented in Figure 12 illustrates the relationships between the initial number of nodes (NbrNdInit) and the lifetime parameters (FND, QND, HND, SFND and LND) for different combinations of (α, β). The discussion has been divided into two subsections presented below, namely conclusions and arguments.

1. Conclusions

- The NbrNdInit and the (α, β) combinations: the correlation between the NbrNdInit and the:
  - FND is negative for the combinations (α, β) going from $(0, 1)$ to $(0.5, 0.5)$. This means that when NbrNdInit increases for these (α, β) combinations, the FND decreases. It ends up becoming slightly positive for the combinations (α, β) going from $(0.6, 0.4)$ to $(1, 0)$.
  - QND is positive for all combinations (α, β) except $(0, 1)$. This means that when the NbrNdInit increases for these (α, β) combinations, the QND also increases.
  - HND is positive for all combinations (α, β). This means that when NbrNdInit increases for these (α, β) combinations, the HND also increases.
  - SFND is also positive for all combinations (α, β).
  - LND is positive with the combinations (α, β) going from $(0, 1)$ to $(0.4, 0.6)$ and a negative correlation for the combinations (α, β) going from $(0.5, 0.5)$ to $(1, 0)$.

- The NbrNdInit with (α) and (β) independently: The correlation between the NbrNdInit and the
  - FND becomes less negative when (α) increases and (β) decreases.
  - QND becomes more positive when (α) increases and (β) decreases.
  - HND remains positive but decreases slightly when (α) increases and (β) decreases.
  - SFND becomes more positive when (α) increases and (β) decreases.

- LND becomes less positive then negative when (α) increases and (β) decreases.

2. Arguments

   - At the start of the network, the nodes have enough energy to support long distance data communications if the energy criterion (α) is privileged. But after several rounds of execution, the nodes that will be selected as *CHs* may be far from the *BS*. The antra and inter cluster activity accelerates their depletion.
   - The correlation results for the variation of (β) are the opposite of those for (α). Privileging the *DistBS* criterion over the energy criterion creates hotspots and a progressive failure of the network starting from the nodes nearest to the nodes furthest from the *BS*. This gives rise to a strong positive correlation between the SFND and LND parameters.
   - The correlation results for the variation of (α, β) which are close to $(0.5, 0.5)$ and which mean that the energy criterion and the *DistBS* criterion share more or less equally the weights, privilege at the same time the nodes which make a compromise between the two criteria in consideration for their election as *CHs*. These combinations optimise energy consumption and increase the lifetime of the network.

5. Vector of Performances Indicators

In this section, a new formal notation called Vector of Performance Indicators (VPI) is presented, which is used to evaluate and measure the performance of a sensitive and energy limited IoT network. The VPI is represented as a set of indicators synthesised from the graph of node mortality as a function of rounds. This set of parameters constitutes a vector of 07 positions defined below. The initial definition of two main properties of the *VPI* forms the basis of the other parameters.

- The JuMping Point (JMP) which indicates the round where the node loss has reached a certain percentage $D1\%$ (e.g. TwPND:Twinty Percent Nodes Dead or TePND: Ten Percent Nodes Dead) which is chosen according to the sensitivity of the application to node loss as well as its functioning. The JMP is considered as a jump point in the mortality graph.
- The DRop Point (DRP), which is the indicator of the round in which the mortality graph dropped after the JMP, i.e. the round in which the network lost a significant amount (percentage) of $D2\%$ nodes (e.g. 80% or 90% of lost nodes). The DRP is also designated according to the application and its objectives, as well as the stability of the network once the network has lost a significant number of nodes.

It is very important to note that the comparison of two protocols is done under the same conditions and assumptions. The *VPI* of a Network i (Nt_i) is symbolized $VPI(Nt_i)$=(FND, FPND, DRP, LND, FLR, SLR, TLR). If the value of the element (j) of the vector *VPI* needs to be specified, it is mentioned as $VPI(Nt_i, j)$. All the parameters of the *VPI* are well explained hereafter as well as schematised in Figure 13.

- $VPI(Nt_i, 1)$: FND (First Node Dead), the round of the first node dead.
- $VPI(Nt_i, 2)$: FPND (forty percent Nodes Dead(40%)), the round where the network has lost forty percent of the nodes.
- $VPI(Nt_i, 3)$: DRP (NPND (Ninety percent Nodes Dead(90%)), the round where the network lost ninety percent of the nodes.
- $VPI(Nt_i, 4)$: LND (Last node Dead), specifies the round of the last node dead.

.

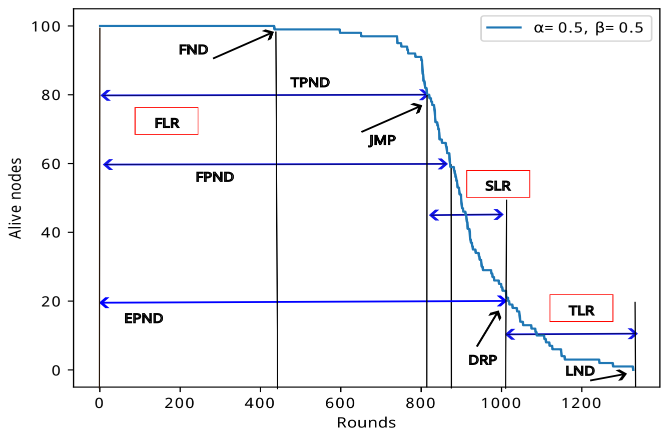


Figure 13: Illustration of the VPI parameters: *DCOPA* $(\alpha = 0.5,\ \beta = 0.5),\ D1 = 20\%\ \ and\ \ D2 = 80\%$

- $VPI(Nt_i, 5)$: FLR (First Level Resistance), this indicator measures the performance of the network resisted until the JMP round where the network loses 10% (TePND) of its nodes.

- $VPI(Nt_i, 6)$: SLR (Second Level Resistence), it is the extent of the interval between the JMP and the DRP.

- $VPI(Nt_i, 7)$: TLR (Third Level Resistence), it is the extent of the interval between the DRP and LND.

This performance formalisation, represented as a vector of integer values, is composed of three resistance levels and four round numbers that focus on very interesting metrics of an energy sensitive IoT network, namely the first and last node dead, the JMP and DRP. This representation allows us, firstly, to compare any two protocols applied to the same network conditions, in terms of lifetime and death rate, and secondly, to indicate the weak and strong points of a protocol by making reference to the different parameters of the *VPI*. In a global way, when developing a clustering and communication protocol in IoT with energy and lifetime optimisation, it is necessary to control and maximise all these parameters of the *VPI*.

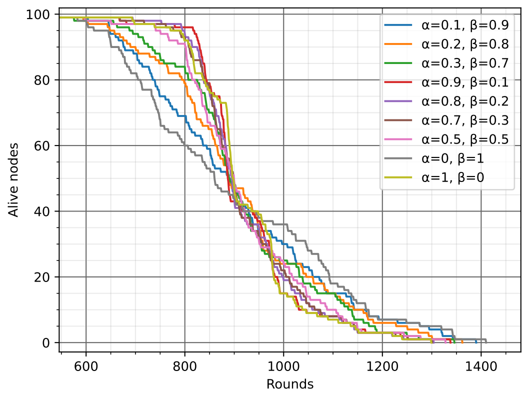
6. *ADCOPA* : Adaptive-*DCOPA*

6.1. The VPI *of* DCOPA *by varying* (α, β)

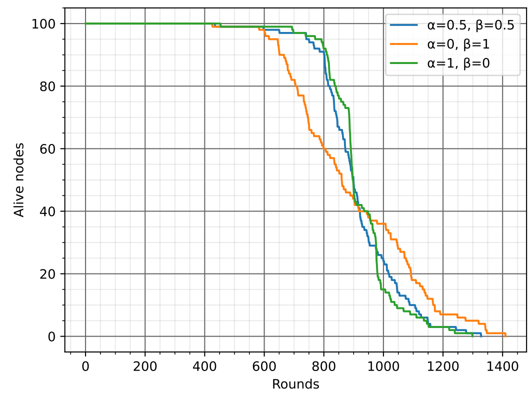
The *DCOPA* protocol was simulated using MATLAB with various static (α, β) combinations. The network configuration presented in the Table 3, the assumptions that were made and the power model that was presented, section 3.1, are similar to the simulation conditions of *DCOPA*. An in-depth analysis was carried out on the influence of the weighting of the two criteria used. The performances are illustrated in Figure ($14a$), and the $VPIs$ of each combination were obtained, which are listed in Table 9. The value of ($D1 = 10\%$) and ($D2 = 90\%$) was chosen to specify JMP and DRP, respectively. Our objective is to make findings on the mortality or degree of loss of nodes using a comparison based on the seven parameters *VPI* for the set of simulated static combinations. The assumptions are listed below.

- The *BS* has an unlimited energy supply.
- The nodes remain stationary.
- The nodes do not possess the equipment to determine their own positions.
- The batteries within the nodes are not replaceable.
- A node will only fail if its energy is entirely depleted.
- All nodes have the ability to adjust their ranges based on their distance from the receiver(s).

After these simulations with static (α, β) combinations and an in-depth analysis of the graphs of Figure ($14a$) and specifically Figure ($14b$) which shows the two monocriteria cases and the one of equal weights for more clarity. Interesting observations and conclusions were drawn concerning the influence of the weights



(a)



(b)

Figure 14: Percentage Nodes Dead as a function of rounds

assigned to energy and *DistBS* on lifetime, mortality rate and the three levels of resistance defined in the *VPI* parameters. Two monocriteria cases, ($\alpha = 1$, $\beta = 0$) and ($\alpha = 0$, $\beta = 1$), were considered, along with several multicriteria cases that will be evaluated in the following.

- ($\alpha = 1, \beta = 0$): the network shows a performance in the FLR metric. After the JMP, the mortality is very significant, i.e., in a few rounds the network has lost a large number of nodes so that the combination has not obtained a satisfactory score in the SLR metric. From the DRP onwards, the mortality stabilises until the last node dead (LND). The LND metric is the lowest of all considered combinations. Contrary to the FLR value, the network in this case shows a more low TLR compared to others.
  The *VPI* of the network, by applying this combination, is presented in vector below and also illustrated in Table 7.
  VPI(N) = (454, 890, 980, 1299, 837, 143, 319).

  Discussion 6: This is due to the fact that this combination is firstly

Table 7: VPI obtained with static weights $(1, 0)$

	Lifetime Parameters				Resistance Levels		
(α, β)	FND	FPND	DRP	LND	FLR	SLR	TLR
$(1, 0)$	454	890	1038	1299	810	228	261

monocriteria where there is just the consideration of energy as the only decisive factor for a node to proclaim itself as *CH*. The consequences described above are due to the fact that nodes with more energy take on the role of *CHs*, which increases the resistance at the first level (FLR) because all nodes have enough energy when the network is launched. At a certain (lower) energy threshold and not checking the *DistBS* of some nodes in the *CHs* choice condition, some nodes far away from the *BS* can be chosen for the role of the *CH*. When aggregating and sending data to the *BS* over a very long distance as well as sending and receiving cluster join messages, this causes their failures which gives rise to a fast drop in the network and thus a less satisfactory result in terms of the two levels of resistance, namely SLR and TLR. As a result, a very early LND relative to the other cases studied will be observed.

- $(\alpha = 0,\ \beta = 1)$: in this case, the network starts losing nodes very early and resists more for the last nodes that will dissipate all their energy. A better case of LND of all combinations was noticed. Very early, the JMP parameter will be obtained, while the DRP parameter will be identified later, enabling a better SLR to be obtained. After the DRP, the network soon goes to extinction, which results in a lower TLR metric.

  The *VPI* of the network, by applying this combination, is presented in vector below and also illustrated in the Table 8.

  VPI(N)=(426, 799, 1192, 1410, 705, 487, 218).

  Discussion 7: The mortality results observed in this combination, which is also monocriteria by considering just the *DistBS* as the only criterion for self-selection of *CHs*, are due to the fact that the nodes closest to the

Table 8: VPI obtained with static weights $(0, 1)$

	Lifetime Parameters				Resistance Levels		
(α, β)	FND	FPND	DRP	LND	FLR	SLR	TLR
$(0, 1)$	426	799	1167	1410	651	516	243

DistBS act as *CHs* several times over very short periods. This makes these nodes hot spots with high communication activity in the network. As a result, their failure is accelerated. That is, very soon, the first nodes dissipating all their energy appear as well as the appearance of the network's drop point, namely the JMP, thus a first level of resistance (FLR) that is less efficient. In this way, the mortality of more and more nodes close to the *BS* gradually increases. This implies a more advanced DRP, a second level of resistance and a more consistent LND.

- ($\alpha \neq 0$ and $\beta \neq 0$): the combinations studied are listed in the column (α, β) of the Table 9. The performances *VPI* are between those of the two monocriteria cases previously mentioned. If the value of (α) is increased, the *VPI* performances tend to approximate those of the case where ($\alpha = 1, \beta = 0$). On the other hand, if the value of (β) is increased, the *VPI* performances tend to approximate those of the case where ($\alpha = 0, \beta = 1$). In terms of the metrics of the vector *VPI*, the combination ($\alpha = 0.5$, $\beta = 0.5$), obtained very satisfactory results on the three levels of resistance FLR, SLR and TLR as well as the metrics FND and LND which qualified as average. This strongly justifies the authors' choice in *DCOPA* (Mir & Meziane, 2023).

  Discussion 8: the performance of the *VPI* retained during the execution of the combinations ($\alpha \neq 0$ and $\beta \neq 0$) highlighting the multi-criteria aspect of energy and *DistBS* for the designation of the *CHs*, is due to the fact of avoiding the considerations resulting from the monocriteria cases, i.e. the hotspots in the case ($\alpha = 0$, $\beta = 1$) and the more distant nodes of the *BS* which win the competition because of their consequent energies

Table 9: *VPI* obtained with different static weights (100 nodes)

	Lifetime Parameters				Resistance Levels		
(α, β)	FND	FPND	DRP	LND	FLR	SLR	TLR
$(1, 0)$	454	890	1038	1299	810	228	261
$(0.9, 0.1)$	412	882	1031	1338	830	201	307
$(0.8, 0.2)$	454	882	1058	1304	815	243	246
$(0.7, 0.3)$	443	875	1063	1301	808	255	238
$(0.6, 0.4)$	438	874	1081	1323	802	279	242
$(0.5, 0.5)$	435	872	1088	1328	801	287	240
$(0.4, 0.6)$	423	872	1110	1332	755	355	222
$(0.3, 0.7)$	420	874	1131	1346	726	405	215
$(0.2, 0.8)$	428	862	1142	1362	690	452	220
$(0.1, 0.9)$	430	837	1143	1390	676	467	247
$(0, 1)$	426	799	1167	1410	651	516	243

in the case ($\alpha = 1,\ \beta = 0$) and which lead to their failure because of the long distances of communication with the *BS*. By avoiding these two disadvantages, the nodes that are likely to win the competition for the role of *CHs* will be those that are closer to the *BS* and have at the same time enough energy. This means that these combinations avoid the extremities in terms of *VPI* of the two monocriteria cases by obtaining results with levels of resistance and lifetime that are close to the average of the two monocriteria cases.

6.2. *Study of the linear correlation of VPI parameters*

In what follows, the linear correlation will be studied between the parameters *VPI* in order to understand the influence of one parameter on another through the definition of the linear correlation coefficients. Figure 15 represents the calculated linear correlation coefficients of the parameters *VPI* of the cases and the simulations values given in Table 9. The main points are:

- (α) influences positively (strong linear correlation (r value close to $+1$)) and (β) influences negatively (strong linear correlation (r value close to -1)) the parameters *VPI* :

	alpha	*beta*	*FND*	*FPND*	*DRP*	*LND*	*FLR*	*SLR*	*TLR*
alpha	1								
beta	-1	1							
FND	0.465711	-0.46571	1						
FPND	0.813583	-0.81358	0.299374	1					
DRP	-0.82298	0.822979	-0.18348	-0.8274	1				
LND	-0.86921	0.869208	-0.58171	-0.90054	0.769052	1			
FLR	0.857936	-0.85794	0.260338	0.875936	-0.94429	-0.87343	1		
SLR	-0.85474	0.854739	-0.23068	-0.86711	0.981553	0.840388	-0.98979	1	
TLR	0.270703	-0.2707	-0.39313	0.240989	-0.66458	-0.03348	0.455453	-0.55234	1

Figure 15: Correlation Matrix of the VPI parameters

  - $FPND$
  - FLR

- (α) influences negatively (strong linear association (r value close to -1)) and (β) influences positively (strong linear association (r value close to $+1$)) the parameters VPI:
  - DRP
  - LND
  - SLR

For the other two parameters FND and TLR, the observed r value was close to 0, this indicates a weak association. As long as the value of (α) is decremented for less interest for the energy criterion and increment (β) for more interest for the criterion *DistBS*, as shown in the Table 9, Observations reveal a continuous improvement in the results for the two lifetime parameters ($VPI(Nt_i, 3)$=DRP) and ($VPI(Nt_i, 4)$=LND). A relative stability in the parameter ($VPI(Nt_i, 1)$ = FND) was observed. For the parameter ($VPI(Nt_i, 2)$ = FPND) the influence is negative. For the resistance level parameters, the influence is negative for the parameter ($VPI(Nt_i, 5)$=FLR) and positive for the parameters $VPI(Nt_i, 6)$=SLR) and ($VPI(Nt_i, 7)$=TLR).

6.3. ADCOPA *: Description and performance evaluation*

It would be unthinkable to conclude our article without mentioning some of the drawbacks of *DCOPA*, after having carried out numerous simulations, analyses, comments, performance evaluations, scalability tests, load balancing, distribution of *CHs* and clusters, as well as the various correlations established between the different performance metrics. These drawbacks are being addressed in order to improve the performance of the protocol in terms of lifetime and energy optimisation. It is therefore crucial to take them into consideration to achieve this goal. In what follows, the discussion and description of certain behaviours of *DCOPA* that are considered weak points for some combinations of (α, β) will be presented. Then, modifications to *DCOPA* are proposed, particularly in the Formula *T(i)* described in Formula 13, precisely at the level of the choice of the weights of the participating criteria in the selection function of *CHs*. This is done in order to obtain better performance in terms of energy management of the entire network and of the nodes, which is reflected in better lifetimes. To this end, dynamic combinations of weights that are sensitive to both the global energy state of the network and the desired performances are implemented (see Formulas 26 and 27).

$$T(i) = \begin{cases} (\alpha_j E_i + \beta_j D_i)(\tau - \delta) & if\ i \in G \\ \tau - \delta & \text{otherwise} \\ \alpha_j + \beta_j = 1 \end{cases} \tag{26}$$

E_i and D_i are defined and detailed in the equations 13, 15 and 16.

$$(\alpha_j, \beta_j) = \begin{cases} (\alpha_1, 1-\alpha_1) & if\ r < RNW_1 \\ (\alpha_2, 1-\alpha_2) & if\ r \geq RNW_1 \ \ \& \ \ r < RNW_2 \\ ... \\ (\alpha_j, 1-\alpha_j) & if\ r \geq RNW_{j-1} \ \ \& \ \ r < RNW_j \\ ... \\ (\alpha_n, 1-\alpha_n) & if\ r \geq RNW_n \end{cases} \tag{27}$$

The combination $(\alpha = 0.5, \beta = 0.5)$ of the *DCOPA* protocol is selected as the optimal choice in terms of *VPI* for static combinations. This indicates our preference for this combination as a starting point for our improvement attempts. The previous discussions in section 6.1 and the in-depth analysis carried out have identified several results. Firstly, the parameters of *VPI* were analyzed and the results were presented in Table 9, showcasing the variation of the studied parameters. Additionally, a study on the linear correlation between these different *VPI* parameters was conducted, as illustrated in Figure 15. Secondly, regarding the global energy consumption results, an analysis was conducted to understand the behavior of node mortality in the network as the combinations of criteria weights changed. The objective is to determine the optimal weight allocation for energy and *DistBS* to enhance the network's lifetime. The basic idea of our contribution is to adapt dynamic combinations of (α, β) in such a way that our solution is context-sensitive and takes advantage of the influence of the criteria weights during specific phases in the life of the network. In our protocol, named *ADCOPA*, the focus will be on a single aspect which is the increase of the lifetime of the network by maintaining a balanced mortality without rushing the network, or simply maximizing the *VPI* parameters defined in the section 5. The question is to identify the values of these combinations as well as the factors to be taken into account for the change of these combination values, which will lead us to our contribution named *ADCOPA* for the improvement of mortality. The main idea is to consider the global energy consumption. Three reference points have been retained during the combination $(\alpha = 0.5, \beta = 0.5)$ in the static case, corresponding to the rounds where the network has dissipated (1/3), (1/2), and (2/3) of its initial global energy. The associated round numbers are respectively 304, 455, and 610, which are named Rounds of New Weights (*RNW*). In the simulations, a single variation is retained for the three *RNWs* to demonstrate the contribution (see Formula 28). It is logical to think of a change of combination from these points of energy dissipation to try to slow down the mortality of the nodes which strongly depends on the energy management of the network and the nodes. The only way to do this is to run several simulations of different combinations based on the successful results of each combination or by comparing to the correlations obtained. The basic idea is that, unlike in *DCOPA* where the combinations are static, *ADCOPA* considers dynamic combinations based on the parameters VPI that are to be optimized.

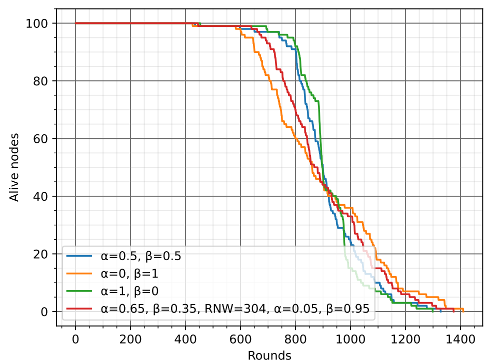
$$(\alpha_j, \beta_j) = \begin{cases} (\alpha_1, 1-\alpha_1) & if\ r < RNW_1 \\ (\alpha_2, 1-\alpha_2) & if\ r \geq RNW_1 \end{cases} \tag{28}$$

There are an unlimited number of combinations of (α, β) to be tested, which can be modified as necessary during the execution of the protocol based on a signal received from the *BS*. A couple of these combinations will be chosen to illustrate our contribution with two cases for each *RNW*. Figure 14 shows the simulated and retained cases that improve the *VPI* parameters presented in Table 10 compared to the static combination of ($\alpha = 0.5$, $\beta = 0.5$).

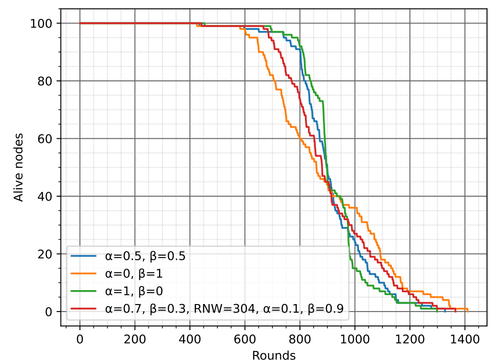
Table 10: VPIs obtained based on the variation of (α, β) weights

			Lifetime Parameters				Resistance Levels		
(α, β) STATUS	RNW	(α, β) VALUES	FND	FPND	DRP	LND	FLR	SLR	TLR
STATIC weights	/	$(0.5, 0.5)$	435	872	1088	1328	801	287	240
DYNAMIC weights	304	$(0.65, 0.35)$,$(0.05, 0.95)$	446	842	1137	1375	722	415	328
	304	$(0.7, 0.3)$,$(0.1, 0.9)$	443	852	1140	1366	723	417	226
	455	$(0.4, 0.6)$,$(0.2, 0.8)$	423	865	1145	1369	707	438	224
	455	$(0.7, 0.3)$,$(0.07, 0.93)$	443	844	1125	1358	751	374	233
	610	$(0.3, 0.7)$,$(0.1, 0.9)$	420	852	1142	1359	714	428	217
	610	$(0.4, 0.6)$,$(0.05, 0.95)$	423	859	1127	1357	738	389	230

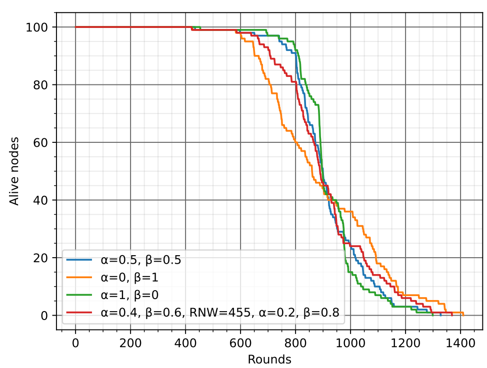
Discussion 9: In this context, the weights of the energy and *DistBS* criteria in the competition of election of the *CHs* will be changed at a given time or in other words at specific rounds which correspond to quite precise levels of energy of the network, named previously *RNW*, to better manage the global energy and the mortality of the nodes. The values obtained from the execution of the dynamic combinations (α, β) are due to the following factors, firstly, drawn very important conclusions in the variations of weights of the (static) criteria (9), very useful syntheses retained from the discussions made previously as well as the factors influencing positively or negatively on the performance of the *VPI*, mainly the levels of resistances. Secondly, by applying the changes in the



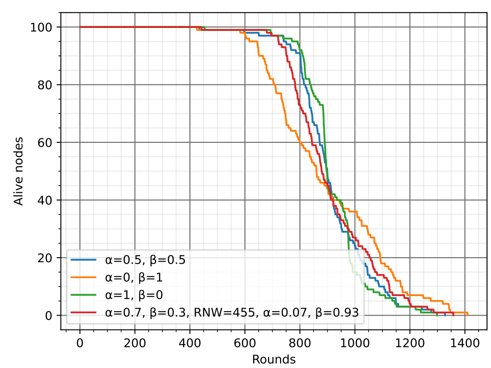
(a) Static weights compared dynamic weights $(0.65, 0.35)$, $RNW = 304$, $(0.05, 0.9)$



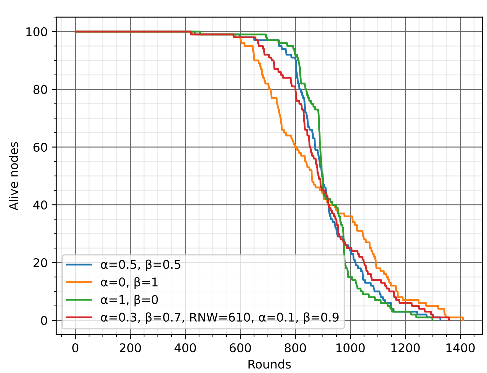
(b) Static weights compared dynamic weights $(0.7, 0.3)$, $RNW = 304$, $(0.1, 0.9)$



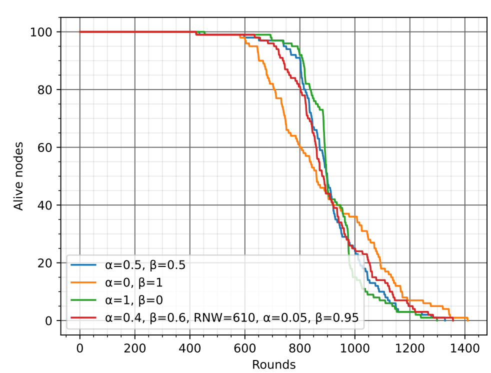
(c) Static weights compared dynamic weights $(0.4, 0.6)$, $RNW = 455$, $(0.2, 0.8)$



(d) Static weights compared dynamic weights $(0.7, 0.3)$, $RNW = 455$, $(0.07, 0.93)$



(e) Static weights compared dynamic weights $(0.3, 0.7)$, $RNW = 610$, $(0.1, 0.9)$



(f) Static weights compared dynamic weights $(0.4, 0.6)$, $RNW = 610$, $(0.05, 0.95)$

Figure 16: Number of alive nodes as a function of rounds : static weights compared dynamic weights

weights of the two criteria when a certain amount (percentage) of energy is dissipated from the network, the dynamicity has been applied in its simplest form, i.e. a single change as indicated in the Formula 28. At the beginning, the energy criterion was heavily weighted ($\alpha > \beta$) because the nodes have their full energies, which influenced the FLR by avoiding hot spots and their premature failure. The defined combination does not hold until the whole network fails. But, at the RNW point (the round which corresponds to the dissipation of (1/3) or (1/2) or (2/3) of the global energy of the network), the *BS* sends a control message to change the weights of the criteria to give more interest to the *DistBS* (α, β), because the nodes have lost a quantity of energy. This means that the nodes which are closer to the *BS* and which have a little more energy can take on the role of the *CH* to try to improve the other two levels of resistances which are SLR and TLR.

7. Conclusion

In conclusion, the research questions raised at the outset have been successfully addressed. Based on an in-depth examination of the DCOPA protocol and its multi-criteria aggregation function, the impact of variations in static weights on *CHs* selection, load balancing, scalability, and cluster distribution has been thoroughly investigated. An improved DCOPA protocol, called ADCOPA, was introduced, using dynamic weighting techniques to optimise energy and mortality rates. *ADCOPA* is a distributed clustering algorithm for data communications in IoT networks. The timer *T(i)*, designed on the basis of a weighted sum which is a multi-criteria optimisation technique, allows a node to access the role of *CH* after a competition between the nodes of the network at the start of each round. The energy of the node and its *DistBS* are the two main criteria of this function. They are associated with predefined weights according to the interest that attributed to them. They depend on the objective and the type of the IoT application. The nodes with a minimal *T(i)* will have more chance to be *CHs* of the current round. *ADCOPA* introduced a new concept of dynamic weighting of the criteria in the multicriteria function *(T(i))* by using a weighted sum. This dynamicity is managed by the *BS* in order to improve the performance of the network, if ever an anomaly is caused by the combination of the previous weights, when a new round is launched, the *BS* changes the combinations of weights which will be communicated to all the nodes. This can be regarded as a sensitivity of the criteria weights to the network context, in particular to energy management and mortality. A formalism called *VPI* has been developed, which takes the form of a seven-position vector. This formalism allows to characterize the performances of a clustering and data communication protocol dedicated to IoT applications. One of the key aspects of this formalism is its ability to measure the rate of node loss, or mortality, over the lifetime of a given protocol. The *VPI* consists of three levels of resistance and four metrics of lifetime. The results of the simulation show that $ADCOPA$ performs well in *VPI* compared to *DCOPA*. These performances, due to the dynamicity observed when changing the weights of the criteria influencing the choosing of the *CHs*, are explained by the fact of initially privileging the nodes which are close to the *BS* for the role of the *CH*. Once arrived at the RNW, nodes with a considerable amount of energy are given priority for the role of the *CH* contrary to *DCOPA* which operates on the basis of fixed weightings of criteria. In the future, the plan is to explore the impact of the position of the *BS* and consider dynamic variations of the (α, β) combinations on multiple levels or as needed during the lifetime of the network.

References

Abdulzahra, A. M. K., Al-Qurabat, A. K. M., & Abdulzahra, S. A. (2023). Optimizing energy consumption in wsn-based iot using unequal clustering and sleep scheduling methods. *Internet of Things*, *22*, 100765. https://doi.org/10.1016/j.iot.2023.100765

Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: a survey. *Computer networks*, *38*(4), 393–422. https://doi.org/10.1016/S1389-1286(01)00302-4

Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. (2015). Internet of things: A survey on enabling technologies, protocols, and applications. *IEEE communications surveys & tutorials*, *17*(4), 2347–2376. https://doi.org/10.1109/COMST.2015.2444095

Arora, V. K., Sharma, V., & Sachdeva, M. (2016). A survey on leach and other's routing protocols in wireless sensor network. *Optik*, *127*(16), 6590–6600. https://doi.org/10.1016/j.ijleo.2016.04.041

Batra, P. K., & Kant, K. (2016). Leach-mac: a new cluster head selection algorithm for wireless sensor networks. *Wireless Networks*, *22*, 49–60. https://doi.org/10.1007/s11276-015-0951-y

Behera, T. M., Mohapatra, S. K., Samal, U. C., Khan, M. S., Daneshmand, M., & Gandomi, A. H. (2019). Residual energy-based cluster-head selection in wsns for iot application. *IEEE Internet of Things Journal*, *6*(3), 5132–5139. https://doi.org/10.1109/JIOT.2019.2897119

Beiranvand, Z., Patooghy, A., & Fazeli, M. (2013). I-leach: An efficient routing algorithm to improve performance & to reduce energy consumption in wireless sensor networks. In *The 5th Conference on Information and Knowledge Technology*, (pp. 13–18). IEEE. https://doi.org/10.1109/IKT.2013.6620030

Benelhouri, A., Idrissi-Saba, H., & Antari, J. (2023). An evolutionary routing protocol for load balancing and qos enhancement in iot enabled heterogeneous wsns. *Simulation Modelling Practice and Theory*, *124*, 102729. https://doi.org/10.1016/j.simpat.2023.102729

Cao, B., Zhang, Y., Zhao, J., Liu, X., Skonieczny, Ł., & Lv, Z. (2021). Recommendation based on large-scale many-objective optimization for the intelligent internet of things system. *IEEE Internet of Things Journal*, *9*(16), 15030–15038. https://doi.org/10.1109/JIOT.2021.3104661

Chang, L., Li, F., Niu, X., & Zhu, J. (2022). On an improved clustering algorithm based on node density for wsn routing protocol. *Cluster Computing*, *25*(4), 3005–3017. https://doi.org/10.1007/s10586-022-03544-z

Chen, H., Zhang, C., Zong, X., & Wang, C. (2013). Leach-g: an optimal cluster-heads selection algorithm based on leach. *Journal of Software*, *8*(10), 2660–2667. https://doi.org/10.4304/jsw.8.10.2660-2667

Dietrich, I., & Dressler, F. (2009). On the lifetime of wireless sensor networks. *ACM Transactions on Sensor Networks (TOSN)*, *5*(1), 1–39. https://doi.org/10.1145/1464420.1464425

Hani, R. M. B., & Ijjeh, A. A. (2013). A survey on leach-based energy aware protocols for wireless sensor networks. *Journal of Communications*, *8*(3), 192–206. https://doi.org/10.12720/jcm

Heinzelman, W. B., Chandrakasan, A. P., & Balakrishnan, H. (2002). An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on wireless communications*, *1*(4), 660–670. https://doi.org/10.1109/TWC.2002.804190

Heinzelman, W. R., Chandrakasan, A., & Balakrishnan, H. (2000). Energy-efficient communication protocol for wireless microsensor networks. In *Proceedings of the 33rd annual Hawaii international conference on system sciences*, (pp. 10–pp). IEEE. https://doi.org/10.1109/HICSS.2000.926982

Hosseinzadeh, M., Hemmati, A., & Rahmani, A. M. (2022). Clustering for smart cities in the internet of things: a review. *Cluster Computing*, *25*(6), 4097–4127. https://doi.org/10.1007/s10586-022-03646-8

Hussain, S., & Matin, A. W. (2005). Energy efficient hierarchical cluster-based routing for wireless sensor networks. *Jodrey School of Computer Science Acadia University Wolfville, Nova Scotia, Canada, Technical Report*, (pp. 1–33).

Junping, H., Yuhui, J., & Liang, D. (2008). A time-based cluster-head selection algorithm for leach. In *2008 IEEE Symposium on Computers and Commu-*

nications, (pp. 1172–1176). IEEE. https://doi.org/10.1109/ISCC.2008.4625714

Kongsorot, Y., Musikawan, P., Muneesawang, P., & So-In, C. (2022). An enhanced fuzzy-based clustering protocol with an improved shuffled frog leaping algorithm for wsns. *Expert Systems with Applications*, *198*, 116767. https://doi.org/10.1016/j.eswa.2022.116767

Kumar, N., Rani, P., Kumar, V., Verma, P. K., & Koundal, D. (2023). Teeech: Three-tier extended energy efficient clustering hierarchy protocol for heterogeneous wireless sensor network. *Expert Systems with Applications*, *216*, 119448. https://doi.org/10.1016/j.eswa.2022.119448

Li, S., Da Xu, L., & Zhao, S. (2015). The internet of things: a survey. *Information Systems Frontiers*, *17*(2), 243–259. https://doi.org/10.1007/s10796-014-9492-7

Mir, F., & Meziane, F. (2023). Dcopa: a distributed clustering based on objects performances aggregation for hierarchical communications in iot applications. *Cluster Computing*, *26*, 1077–1098. https://doi.org/10.1007/s10586-022-03741-w

Murata, T., & Ishibuchi, H. (1994). Performance evaluation of genetic algorithms for flowshop scheduling problems. In *Proceedings of the First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence*, (pp. 812–817). IEEE. https://doi.org/10.1109/ICEC.1994.349951

Nord, J. H., Koohang, A., & Paliszkiewicz, J. (2019). The internet of things: Review and theoretical framework. *Expert Systems with Applications*, *133*, 97–108. https://doi.org/10.1016/j.eswa.2019.05.014

Rahayu, T. M., Lee, S.-G., & Lee, H.-J. (2014). Survey on leach-based security protocols. In *16th International Conference on Advanced Communication*

Technology, (pp. 304–309). IEEE. https://doi.org/10.1109/ICACT.2014.6778970

Saporta, G. (2006). *Probabilités, analyse des données et statistique*. Editions technip.

Shreyas, J., Chouhan, D., Rao, S. T., Udayaprasad, P., Srinidhi, N., & Kumar, S. D. (2021). An energy efficient optimal path selection technique for iot using genetic algorithm. *International Journal of Intelligent Internet of Things Computing*, *1*(3), 230–248. https://doi.org/10.1504/IJIITC.2021.115705

Singh, S. K., Kumar, P., & Singh, J. P. (2017). A survey on successors of leach protocol. *Ieee Access*, *5*, 4298–4328. https://doi.org/10.1109/ACCESS.2017.2666082

Srinidhi, N., Kumar, S. D., & Venugopal, K. (2019). Network optimizations in the internet of things: A review. *Engineering Science and Technology, an International Journal*, *22*(1), 1–21. https://doi.org/10.1016/j.jestch.2018.09.003

Tong, M., & Tang, M. (2010). Leach-b: an improved leach protocol for wireless sensor network. In *2010 6th international conference on wireless communications networking and mobile computing (WiCOM)*, (pp. 1–4). IEEE. https://doi.org/10.1109/WICOM.2010.5601113

Triantaphyllou, E. (2000). Multi-criteria decision making methods. In *Multi-criteria decision making methods: A comparative study*, (pp. 5–21). Springer. https://doi.org/10.1007/978-1-4757-3157-6_2

Zhao, Z., Li, G., & Xu, M. (2019). An improved algorithm based on leach routing protocol. In *2019 IEEE 19th International Conference on Communication Technology (ICCT)*, (pp. 1248–1251). IEEE. https://doi.org/10.1109/ICCT46805.2019.8947022