This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3092509, IEEE Access

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Towards Energy Efficient Clustering in Wireless Sensor Networks: A Comprehensive Review

NASSIMA MERABTINE[1], **DJAMEL DJENOURI** [2], **DJAMEL-EDDINE ZEGOUR.**[1]
[1]Computer Science Higher School, ESI, ex. INI, Algiers, ALGERIA
[2]Computer Science Research Center, Department of Computer Science and Creative Technologies, University of the West of England, Bristol, UK
Corresponding author: Nassima Merabtine (e-mail: nassimane@gmail.com).

ABSTRACT Clustering is one of the fundamental approaches used to optimize energy consumption in wireless sensor networks. Clustering protocols proposed in the literature can be classified according to different criteria related to their features such as the clustering methodology, objectives, cluster count and size, etc. This paper reviews the existing feature-based classifications of clustering protocols and elaborates a more generic and unified classification. It also analyzes and discusses the relevant design factors that may influence the energy efficiency of clustering protocols and accordingly proposes a new energy-oriented taxonomy. State-of-the-art clustering solutions are then reviewed and evaluated following the proposed taxonomy.

INDEX TERMS Wireless Sensor Networks; Clustering Protocols; Energy Efficiency.

I. INTRODUCTION

A Wireless Sensor Network (WSN) is a collaborative infrastructure composed of a number of tiny, wireless, battery-powered nodes called Sensor Nodes (SNs), and (a) powerful node(s) called Base Station(s) (BS). SNs are resource-constrained, i.e, limited in memory storage, processing capability, bandwidth, and battery. Moreover, SNs are usually deployed in remote, inaccessible and sometimes hostile areas, e.g., in battlefields or borders (military applications), or implanted in human body (healthcare applications). Consequently, it is very difficult, costly and in some cases impossible to recharge or substitute SNs' batteries. For these reasons, optimizing energy consumption and extending the WSN lifespan has been the major concern of the WSN research community over years. Furthermore, a wide range of WSN applications require deployment of sheer number of nodes (hundreds or thousands of SNs) to achieve their ultimate goal. The management of such a large and dense deployment relates to *scalability* [1], which can be defined as the network capability to cope and perform under a large or increased number of nodes while maintaining the network performance. Relying on a flat physical topology cannot ensure scalability as the transmission power is proportional to the transmission range [2] that is likely to be high in a flat topology, in which SNs directly transmit to the BS(s). This is inefficient due to the fact that the transceiver (transmitter/receiver circuit) is the most energy hunger component of a SN [3].

To deal with this problem, researchers used "topology control methods" that construct an optimized virtual topology over the WSN. The most relevant method is clustering [4], [5], which consists in dividing the network into a number of groups (clusters), with a particular node designated as Cluster-Head (CH) for every cluster. CHs collect data from their members and forward it to the BS. The CHs may perform other tasks such as aggregating the collected data or scheduling media access for their members. The key principle in clustering is to localize message transmission within clusters, and between CHs and BS, which has many advantages such as preserving the bandwidth, preventing redundancy, and reducing communication overhead. The CH-nodes election (or selection) method is crucial, since the processing and transmission load will be concentrated at those nodes. Numerous clustering protocols have been proposed thus far [6], and every protocol considers some basic requirements such as coverage and connectivity, while dealing with some WSN challenges that are specific to the application, e.g., ensuring stability in presence of node mobility and/or nodes failure, quality of service and security, etc. Although the considered requirements and challenges differ

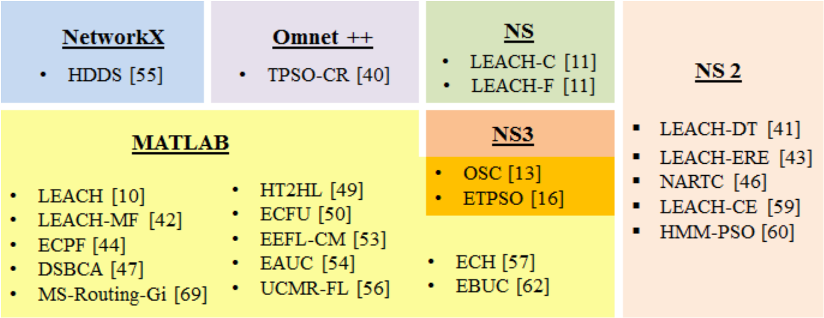

FIGURE 1: Simulation tools used by some state-of-the-art clustering protocols in WSNs.

from a protocol to another, energy-efficiency is a common and shared objective.

The empirical approach has been used in many works to evaluate and compare the proposed clustering protocols with respect to different metrics including network lifetime, number of data signals received at the BS, etc. [7] [8]. To implement the clustering protocols, different network simulators such as NS2, NS3, Omnet++, and network emulators such as TOSSIM have been used. MATLAB has also been used in several works, especially when envisaging analytic evaluation. WSN simulators differ in their design, goals, and characteristics. The selection of the appropriate simulator depends on the requirements and objectives of the solution [9]. Figure 1 illustrates the tools used by some state-of-the-art clustering protocols to validate and compare their performances. It can be noticed from this figure that the majority of the studied protocols (approximately 60%) consider MATLAB in the simulation phase. This is due to its simplicity and the wide range of mathematical libraries it provides. However, this comes at the use of a high level of abstraction ignoring many network/communication aspects and thus reducing accuracy.

On the other hand, many survey papers analyzed and theoretically evaluated existing clustering protocols, which leads to the proposition of several classifications and taxonomies. To the best of our knowledge, all the surveyed papers propose feature-based classifications which consider the protocols' characteristics and properties, such as the method used to choose the CHs, the use of centralized vs. distributed algorithms, etc. This is without reflecting the impact of these characteristics on the energy efficiency of the protocol. The contribution of this paper is threefold: i) Elaboration of a generic and unified feature-based classification of clustering protocols. ii) Proposition of a novel energy-oriented classification that emphasizes the relevant design factors influencing the energy efficiency, and iii) Review and theoretical evaluation of some state-of-the-art and canonical clustering protocols from the perspective of the proposed taxonomy.

The rest of this paper is organized as follows. Section II presents the existing feature-based classifications, followed by the proposed taxonomy of classification attributes in Section III. Section IV discusses the design factors that affect the energy efficiency of clustering solutions and introduces a novel energy-centric taxonomy. SectionV reviews and evaluates some of the existing clustering protocols. Finally, Section VI highlights the important contributions of this paper and draws conclusions. All the terms (abbreviations / notations) used throughout this paper are alphabetically classified in Table 1.

II. RELATED WORK: EXISTING FEATURE-BASED CLASSIFICATIONS

In the following, some general concepts of clustering are presented. Those have been used in numerous survey papers published from 2007 through out 2020. This will allow to elaborate a global and unified feature-based taxonomy of the most relevant clustering attributes in the next section.

The first proposed clustering protocols were static, i.e., the CHs are chosen and the clusters are formed at the network initialization and remain fixed throughout the network lifespan. In this case, the CHs are expected to consume more energy as compared to the cluster members due to their long transmission range to reach the BS and the processing burden to manage their members. This leads to an uneven energy consumption in the network and short network lifetime. To address this issue, Heinzelman et al. introduced the concept of dynamic clustering by proposing the LEACH protocol [10]. LEACH suggests that the role of CH should be rotated periodically among the SNs to ensure a fair distribution of the load and hence a fair distribution of energy consumption in the network. The periodic rotation of the CH role is known also as re-clustering process. Thenceforth, all the efforts have been concentrated on proposing new and efficient dynamic clustering solutions.

Dynamic cluster-based protocols perform on rounds, each one consists of two phases: (1) set-up phase, and (2) steady state phase. In the set-up phase, the clustering establishment process is carried out by exchanging control messages containing some information such as the node ID, the energy level, etc. The set-up phase contains two main steps: (a) CHs election and (b) clusters formation. In the CHs election step, some geospatial and/or energy related criteria are considered to designate the appropriate SNs as CHs for the incoming

Term	Description
AB	Attribute Based
BS	Base Station
CECP	Centralized Evolutionary Clustering Protocol
CH	Cluster-Head
CH times	how many times the SN has been elected as CH
d(SN,BS)	distance between SN and BS
D-LEACH-F	Dynamic round-time Based on LEACH-F
DSBCA	Distributed Self-organization Balanced Clustering Algorithm
DVB	Direct Virtual Backbone
E	Energy
EAUCA	Energy-Aware Unequal Clustering Algorithm
EBUC	Energy-Balanced Unequal Clustering
ECFU	Energy based Clustering with Fuzzified Updates
ECPF	Energy-aware distributed Clustering Protocol using Fuzzy logic
ECH	Enhanced Clustering Hierarchy
EEFL-CH	Energy Efficient Fuzzy Logic Cluster Head
ETPSO-CR	Enhanced Two-tier Particle Swarm Optimization protocol
FCM	Fuzzy C-Means
FL	Fuzzy Logic
FND	First Node Death
GA	Genetic Algorithm
HDDS	Hierarchical Data Dissemination Strategy
HL-FCM	Hybrid LEACH and Fuzzy C-Means algorithm
HMM	Hidden Markov Model
HMM-PSO	Hidden Markov Model and Particle Swarm Optimization protocol
HSA	Harmony Search Algorithm
HSACP	Harmony Search Algorithm-based Clustering Protocol
HT2HL	Hybrid Threshold sensitive and two-Level Heterogeneous
ICD	Intra-Cluster Distance
LEACH	Low Energy Adaptive Clustering Hierarchy
LEACH-C	Centralized LEACH
LEACH-CE	efficient clustering through Estimate in Centralized LEACH
LEACH-DT	LEACH-Distance-based Thresholds
LEACH-ERE	LEACH-Expected Residual Energy
LEACH-F	Fixed LEACH
LEACH-MF	LEACH-Mobile-Fuzzy
IEEHCS	Improved Energy Efficient Hybrid Clustering Scheme
LPO	Lion Pride Optimizer
LPOBC	energy-aware Clustering method via Lion Pride Optimizer and FL
MMDCP	Multilevel Minimized Delay Clustering Protocol
MOFCA	Multi-Objective Fuzzy Clustering Algorithm
MS-routing-Gi	routing technique to minimize energy consumption and packet loss in WSNs with Mobile Sink
NARTC	Network Adaptive Round-Time Clustering
ND	Node Degree
NP-hard	Non-Polynomial-hard problem
OSC	One-Step Clustering protocol
PSO	Particle Swam Optimization algorithm
RF	Radio-Frequency
RSSI	Received Signal Strength Indicator
SA	Simulated Annealing
SN	Sensor Node
TDMA	Time Division Multiple Access
TPSO-CR	Two-tier Particle Swarm Optimization protocol for Clustering and Routing
TTDFP	Two-Tier Distributed Fuzzy logic based Protocol
UCMR-FL	Unequally Clustered Multi-hop Routing protocol based on FL
WSN	Wireless Sensor Network

TABLE 1: Nomenclature of the used terms.

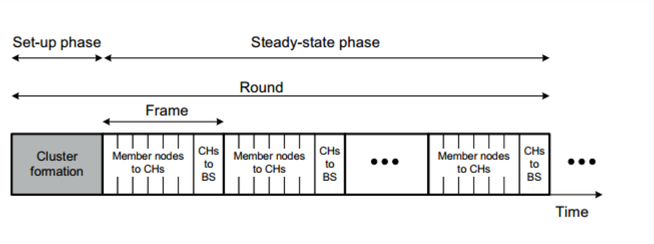

FIGURE 2: Some operation phases of dynamic clustering protocols in WSNs.

round. The non-CH SNs are then assigned to the elected CHs in the cluster formation step. Clustering protocols differ from the way they implement these phases. At the end of the set-up phase, every node knows whether it is a CH or to which cluster it belongs. Thereafter, the network goes into the steady state phase which is divided into a number of frames. In each frame, SNs perform periodical data gathering and send the collected data to their respective CHs, that may make filtering and aggregation before transmitting to the BS. Figure 2 illustrates the general approach of dynamic clustering protocols.

Numerous dynamic clustering protocols have been proposed in the literature, e.g. [11]–[16]. These protocols provide optimization solutions to each phase of the clustering including the CHs election, and the clusters construction. Clustering protocols can be classified according to the way of implementing each clustering phase. Moreover, many other criteria are used, such as the clustering objective, the communication pattern between the member nodes/CHs, and the CHs/BS.

The first comprehensive survey enumerating criteria that can be used to categorize clustering protocols in WSNs was published in 2007 by Abbasi et al. [6]. The taxonomy of criteria proposed in this paper is composed of three parts: 1) cluster properties which include cluster count, stability, intra-cluster topology and inter-CH connectivity, 2) CH capability that includes mobility, node type and role, and 3) clustering process that includes methodology, objective of node grouping, CH selection and algorithm complexity. Several other surveys extended this taxonomy by defining new classification criteria or attributes [17]–[38]. For example, Deosarkar and al. [17] focused on the importance of the CH selection step. They proposed new attributes to classify clustering protocols according to the policy used to pick the CHs which can be deterministic, adaptive, or a combined metric. Jiang et al. [19] added three classification criteria, namely, existence, explicit control messages, and overlapping. Some works use other appellations for some existing classification criteria. For instance, in [19] the authors use the criteria hop distance, selectivity, and count variability, that have been defined in [6] as intra-cluster topology, CH selection, and cluster count, respectively. In the following, we elaborate a global and unified feature-based classification of clustering protocols.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3092509, IEEE Access

III. GLOBAL FEATURE-BASED CLASSIFICATION OF CLUSTERING PROTOCOLS

In this section, we enumerate and discuss the set of the relevant attributes used in the literature to classify and differentiate clustering protocols in WSNs. These attributes are categorized into three groups, 1) clustering process, 2) cluster properties, and 3) CH characteristics. All the studied attributes are organized in a generic taxonomy that is presented in Figure 3.

A. CLUSTERING PROCESS

The design and the characteristics of the overall clustering process differ from a clustering scheme to another. Hereafter, we present the most relevant attributes making these differences.

1) **Methodology:** Three methods for clustering in WSNs can be distinguished: 1) centralized, 2) distributed, and 3) hybrid. In the first one, the BS is responsible for the network clustering (construction of clusters and selection of respective CHs). Therefore, it should have complete information of the network (location of SNs, their battery levels, etc.). In the contrary, no central control exists in distributed (decentralized) clustering, but all the SNs are involved in the clustering process. They execute distributed algorithms and collaborate to select CHs and to form clusters. The third approach is the combination of the two methods and is generally used when some SNs in the network are resources-rich. For instance, the designation of CHs may be done by the BS, while the CHs collaborate to form their clusters in a distributed way.
2) **Clustering approach:** According to the approach used to elect CHs and form clusters, clustering solutions can be categorized as 1) simple-model-based, 2) meta-heuristic-based, 3) fuzzy-logic-based, and 4) hybrid. In the first approach, clustering protocols use a simple formula to elect CHs considering one or more criteria. However, since clustering is an NP-hard problem, some clustering solutions (those of the second approach) benefit from meta-heuristic algorithms to form efficient clusters. Another category of clustering protocols are based on a FL for clustering and selecting the appropriate CHs. This is because the parameters affecting the role of CH may be overlapping. To take the advantage from FL and meta-heuristic algorithms, a few recent solutions (hybrid) propose to combine the both.
3) **Objective of node grouping:** Depending on the targeted application, several objectives have been pursued in the literature including: network lifetime extension, scalability, data aggregation/fusion, fault-tolerance, load balancing, network connectivity, quality of service, etc. Clustering protocols in WSNs can be classified according to the objective of node grouping. A clustering algorithm has usually more than one objective, while the "network lifetime extension" remains the most common objective. Hereafter, we summarize some relevant clustering objectives.

   - Network lifetime: Protocols attempt to reduce the total energy dissipated in the network, which systematically increases the network lifetime.
   - Scalability: Depending on the application, the number of deployed SNs in the sensing zone can be in the order of hundreds, thousands or even more. As mentioned before, clustering the network can ensure the scalability by localizing messages transmission, minimizing the number of messages circulating in the network, etc.
   - Data Aggregation/Fusion: In dense deployment of WSNs, the data collected by the nearby SNs can be similar or correlated. Data aggregation/fusion is an effective approach to avoid transmitting repetitive data in the network by combining data from different sources. Clustering allows the data being easily aggregated in the CHs, which reduces number and size of the transmitted data packets.
   - Fault-tolerance: WSNs are usually prone to failures. In real deployment of WSNs, the failure of some nodes should not interrupt the overall functioning of the WSNs. Many clustering protocols operate in rounds and re-cluster the network at the beginning of everyone. This will allow to handle the faulty nodes. Some protocols define approaches to select a backup CH, or deputy-CH, to cover faulty CHs.
   - Load balancing: It reflects how evenly the energy consumption is distributed among the SNs. Many clustering protocols focus on balancing the roles to avoid that some SNs dissipate energy rapidly compared to the rest of the SNs in the network.
   - Network stabilization: Managing the changes in the network topology is generally more convenient in clustered WSNs than in flat architecture. This is because the SNs are grouped in clusters. Consequently, the changes can be easily detected and handled by the CHs if a SN runs out of battery or moves to other clusters (in a mobile network).
   - Connectivity: Clustering the SNs can facilitate and improve the network connectivity, especially in large-scale deployments of WSNs. A single path from every CH to the BS is sufficient to achieve the connectivity in clustered networks. For the applications requiring connectivity between every couple of nodes, a path between each couple of CHs can be used to assure this.
   - Utilizing sleeping schemes: In some applications of WSNs, there is no need for the nodes in the network to be awake in all the operational time. The SNs can thus enter into a sleep mode to preserve their energy. Clustering the WSN enables a soft implementation of this idea.

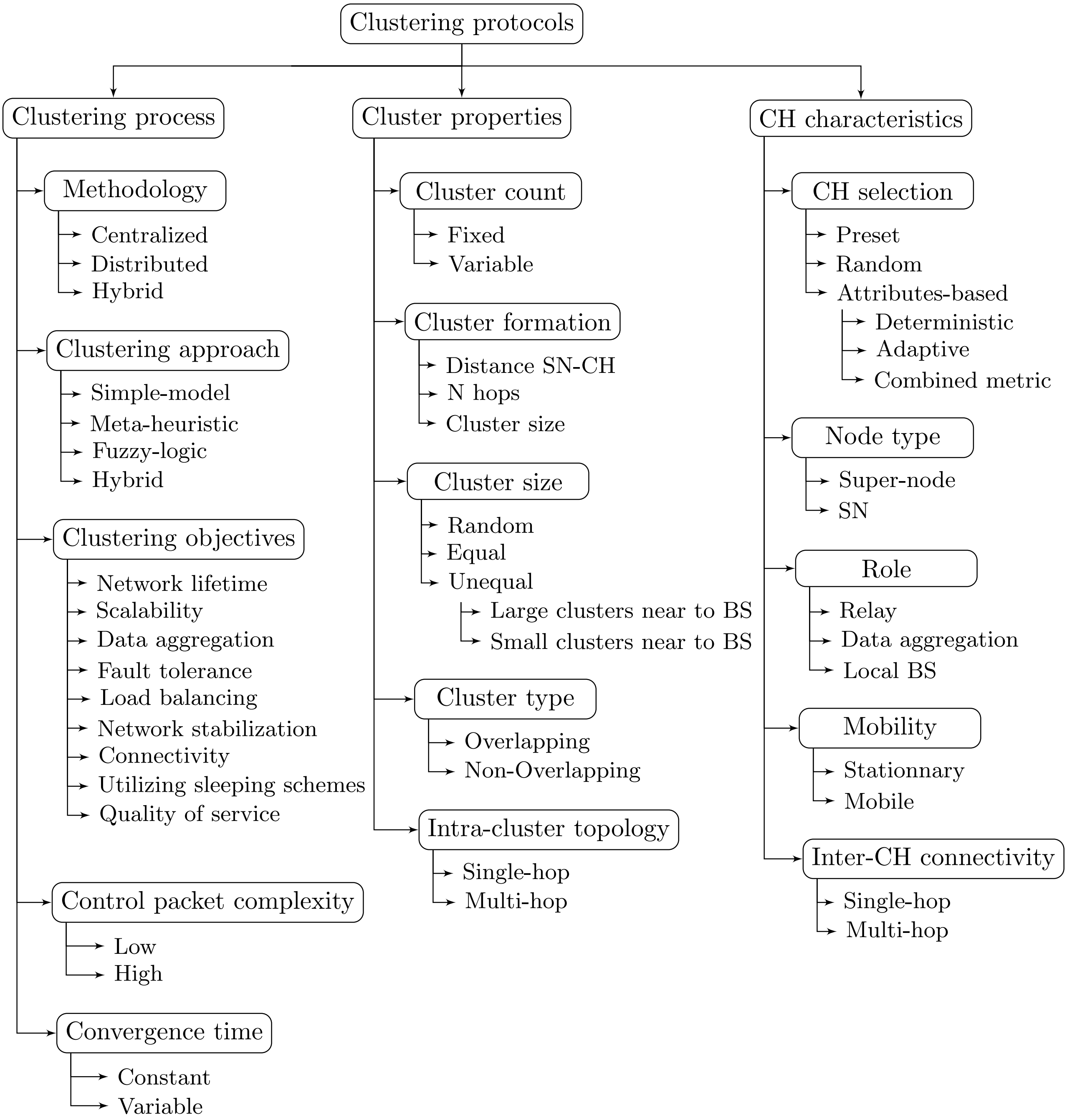

FIGURE 3: Taxonomy of clustering attributes in WSNs.

- Quality of service (QoS): Clustering the network can improve different QoS parameters required by several WSN applications. As previously discussed, clustering protocols minimize the energy dissipation and permit to keep SNs alive for a longer time, which contribute to increase the throughput and enhance network reliability. The letter is also enhanced by increasing the rate of successful data transmission through i) minimizing the interference since only CHs have to communicate with the BS, and ii) avoiding collisions when performing in-cluster operations. Collisions can be straightforwardly avoided if the CHs schedule the access to the shared transmission medium for their members, by using TDMA for example. Furthermore, an efficient clustering solution can minimize the end-to-end delay as transmitting data hierarchically using CHs reduce the number of hops and hence the communication delay.

4) **Control packets complexity:** The complexity of a clustering protocol is measured as a function of the number of control packets exchanged during the clustering process, i.e. to select CHs and form clusters. This is also called control packet overhead of the protocol.
5) **Convergence Time:** This can be constant or variable. In the first case, the number of iterations (thus the number of exchanged control packets) needed for clustering the network is constant independently from the size of the network. In the second case, this number is dependent on the size of the network. Constant convergence time clustering protocols are more scalable.

B. CLUSTER PROPERTIES

Each clustering solution tries to achieve some characteristics for the generated clusters. These characteristics are related to many aspects such as the way the clusters are constructed, the communication paradigm between SNs and their CH, etc. The relevant cluster properties are introduced in the following:

1) **Cluster count:** In heterogeneous WSNs where some resource-rich nodes are deployed to serve as CHs or in some application environments where the set of CHs are predetermined, the number of clusters is preset. On contrary, picking the CHs from the deployed SNs usually yields variable number of clusters. Thus, clustering protocols can be grouped into preset (fixed) cluster-count vs. variable cluster-count.
2) **Cluster construction:** The following factors are used to decide about the admission of an SN in a cluster, or by the SN to select which cluster to join. 1) The distance between the SN and the CH. In centralized clustering protocols, the BS calculates this distance since it has a complete view of topology, while in distributed clustering protocols, the RSSI of the packets sent by CH is used by the SN to calculate this distance. 2) The number of hops between the SN and the CH (in case of multi-hop communication). 3) The desired cluster size (in terms of member). In some cases, a maximum size is defined and the CH may refuse an SN join-request when this threshold is achieved.
3) **Cluster size:** Three categories of clustering protocols can be distinguished here: 1) random, 2) equal, and 3) unequal. In the first category, no importance is given to the cluster size and the SNs join clusters at will. In the second category, the network is divided into clusters of the same cardinality, while to the third category clusters of different sizes are generated. In general, clustering with different sizes is used to achieve load balancing and avoid the energy hole problem. This inequality of clusters' sizes is based on the distance between the nodes and the BS, as well as the communication paradigm between the CHs and the BS. In the case of direct communication (single-hop) between the CHs and the BS, the clusters located at a high distance from the BS should have a smaller size than those close to the BS. Further, when the multi-hop communication is used between the CHs and the BS, the clusters close to the BS should be less overloaded. This is to permit their CHs to save some energy for ensuring routing.
4) **Cluster type:** Traditionally, clustering algorithms aim at generating a number of disjoint clusters that satisfy some criteria. Other clustering protocols utilize overlapping techniques where an SN may belong to more than one cluster. Overlapping clusters are useful in many WSN applications, including node localization, and time synchronization protocols [39]. In summary, clustering schemes can be grouped into overlapping clustering vs. non-overlapping clustering.
5) **Intra-cluster topology:** In some clustering protocols, member nodes communicate their data directly to their respective CHs, while in other protocols, each member node transmits its data to its CH through its neighbor nodes using multi-hop communication. Hence, depending on hop distance between the member node and its CH, clustering schemes can be categorized into single-hop clustering vs. multi-hop clustering.

C. CH CHARACTERISTICS

Clustering protocols can be differentiated according to features related to CHs. The most relevant features are listed below.

1) **CH selection:** This is about the approach and criteria used for CHs selection. Clustering protocols can be categorized into three main classes: 3) pre-assigned, 2) random, 3) attributes-based. In the first class, a subset of SNs in the network is pre-selected to act as CHs. This is practical when the network is heterogeneous. In the second class, the CHs are picked randomly from the deployed nodes. In attributes-based clustering protocols, some criteria are used for CH selection. According to the criteria used, clustering solutions of this class can be further grouped into three subcategories: a) Deterministic: In this subcategory, the inherent attributes of the SN such as its ID, position and the number of neighbors it has (SN degree), are considered. b) Adaptive: in which the resource information drive the CH selection, e.g. residual energy of the SN, energy dissipated during last round, the energy to be dissipated if the SN is chosen as CH, and the communication cost. c) Combined-metric: this subcategory uses a combination of the aforementioned deterministic and adaptive approaches.
2) **CHs type:** The CH can be either a super-node or ordinary node. In heterogeneous WSNs, some super-nodes equipped with significantly more computation and communication resources are deployed in the network. These nodes are generally designated as CHs. On the other hand, CHs are selected from the ordinary SNs in homogeneous WSNs.
3) **Role:** In addition to relaying the data collected by their cluster members to the BS, CHs may perform

some processing such as data aggregation/fusion and then transmit only meaningful data to the BS. A CH may also act as a local BS that takes actions when a phenomenon is detected. Therefore, according to the role of the CH in the network, clustering protocols can be grouped into: relay, aggregation, and local BS.

4) **Mobility:** The CHs can be mobile in some applications. When a CH is mobile, the cluster membership dynamically changes and the clusters would need to be continuously maintained. On the other hand, stationary CH tends to yield stable clusters and facilitate intra and inter-cluster network management.
5) **Inter-CH connectivity:** Similarly to the intra-cluster topology, two types of communication may exist between the CHs and the BS, direct vs. multi-hops. Some of the existing clustering protocols assume that the CHs are able to directly reach the BS. Still, when the CHs do not have long haul communication capabilities or when further energy conservation is targeted, CHs may transmit their data to the BS via multi-hop routes.

IV. ENERGY-ORIENTED TAXONOMY

To measure the energy dissipation of the SNs, clustering protocols make use of energy consumption models. During its life cycle, a SN spends its energy in sensing, processing, and radio communication [5]. There are different radio energy models in the literature. The most common one is the first order radio model [11], in which the energy consumed in transmission, E_{tx}, is proportional to the data size, l, and the transmission distance, d. If d is smaller than a threshold d_0, which is calculated by Eq. 1, the free space model is considered and E_{tx} is calculated using Eq. 2. Otherwise, the multi-path fading career is considered and Eq. 3 is utilized to calculate E_{tx}. The energy dissipated in the receiving mode, E_{rx}, is relative to the data volume only and calculated using Eq. 4. E_{el} is the energy needed to run the transceiver electronic circuitry. ϵ_{fs} and ϵ_{mp} are the amplification energy in the free space and multi-path fading careers, respectively. It is worth noting that some works consider more realistic radio energy models, such as [40] which uses a model based on the CC2420 radio transceiver.

$$d_0 = \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}}, \tag{1}$$

$$E_{tx}(l,d) = lE_{el} + l\epsilon_{fs}d^2, if(d <= d_0), \tag{2}$$

$$E_{tx}(l,d) = lE_{el} + l\epsilon_{mp}d^4, if(d > d_0), \tag{3}$$

$$E_{rx}(l) = lE_{el}, \tag{4}$$

In this section we address the clustering from an energy-efficiency perspective and identify the factors that can impact the energy-effectiveness of a solution. Some of these factors are related to the clustering process such as the clustering methodology, the number of clusters that should be formed, and the size of each formed cluster; while others are related to the clustering protocol characteristics and its capacity to balance the energy dissipation between the different nodes in the network. All the identified factors are discussed in the following and summarized in Figure 4.

A. PROTOCOL DESIGN

1) **Clustering Methodology:** The choice between the distributed and centralized clustering methodologies has an impact. Distributed protocols have some advantages such as self-adaptation, fast execution, and fault tolerance. Nevertheless, centralized protocols are more energy-efficient, as they take advantage of the BS' global view of the network topology and its strong computation capability for executing sophisticated optimization algorithms. Consequently, these protocols produce better clustering quality and balanced clusters while considering the network parameters such as remaining energy of every SN, etc. They also ensure a balanced distribution of the CHs over the network, which results in significant energy saving and network lifetime extension. However, centralized approaches suffer from the communication overhead given that they require SNs to send some information such as residual energy and localization at the beginning of each round. An important amount of energy is therefore dissipated due to this communication overhead. In contrast, only a few messages are exchanged between neighbor nodes in distributed protocols.
2) **CH Selection:** The split of SNs into clusters depends on the number and location of the CHs. Hence, the optimal selection of CHs helps to further reducing energy consumption and extending the network lifetime. Remember that there are three main classes of CHs election methods: pre-assigned, random, and attributes based (Figure 3). The pre-assignment of CHs can be an effective method only if the network is heterogeneous and the selected CHs are super-nodes. If the CHs are selected randomly, two important problems emerge. First, the distribution of the CHs across the network is not performed properly. This may causes concentration of CHs in some regions, and thus long distance between the SNs and their corresponding CHs. Second, the energy of the CHs is not considered in the CH selection, and hence the nodes with low energy may get elected as CHs. Attributes-based methods that consider the energy of SNs, node position to minimize the intra-cluster communication cost, distance to the BS, etc., are more energy-efficient. Furthermore, the number of CHs (which reflects the number of clusters) can also affect the energy efficiency of clustering. A small number of CHs seems to be a good design choice in terms of energy dissipation as it decreases the number of long-range transmission between CHs and the BS (direct or multi-hop), as well as the network contention. However, this can cause long transmission distances between the member nodes and their CHs, which increases the intra-cluster energy consumption and makes the CHs overloaded with more members.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3092509, IEEE Access

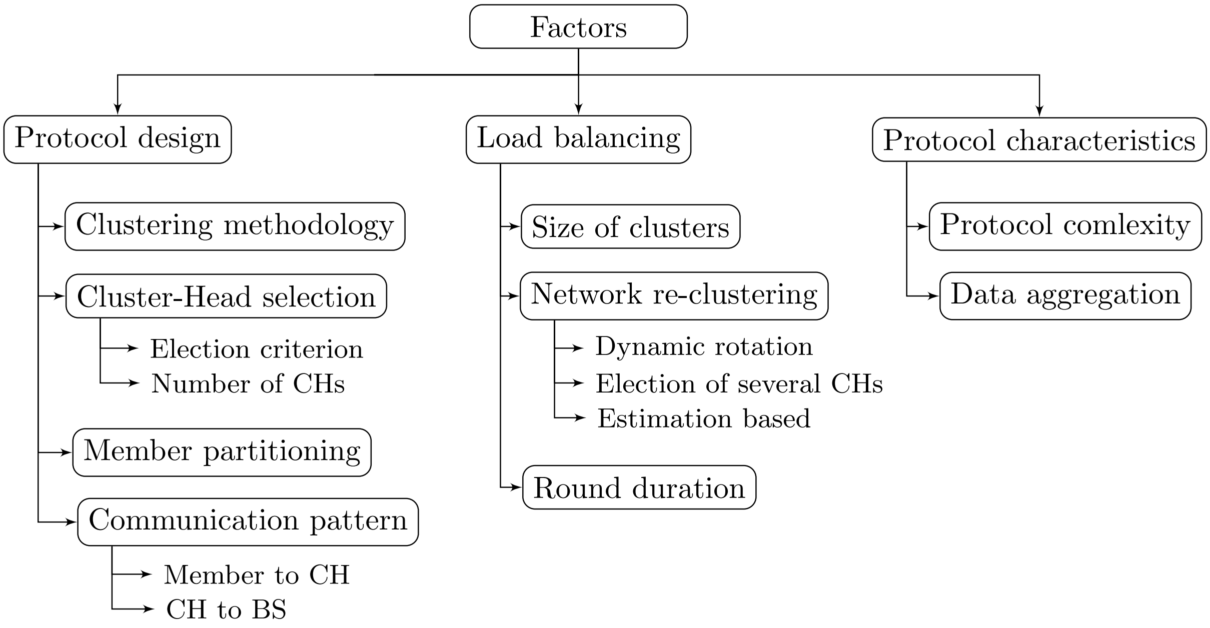

FIGURE 4: Taxonomy of factors influencing the energy efficiency of clustering protocols in WSNs.

Many existing clustering protocols fix this parameter to 5% of the network size, without deep analysis of the energy-effectiveness of such decision. Other solutions use more efficient techniques such as FL to find the appropriate CHs number. Still, finding the optimal number of CHs is an important parameter for optimizing energy usage.

3) **Member Partitioning:** The way members are assigned to CHs can also affect the total energy dissipated in the network. Random assignment of members to CHs is far from being an effective approach. Some techniques such as considering the distance between the member nodes to the CHs is practical to minimize the energy dissipated by the non-CH nodes.
4) **Communication Pattern:** The type of communication between the member nodes and their respective CHs (intra-cluster communication), and between the CHs and the BS (inter-cluster communication) is also an important issue. If the transmission distance is relatively small, which is generally the case in communications between members and CHs, the use of direct communication can be judicious. In contrast, multi-hop communication is preferred and likely to provide a significant energy saving when this distance is high. This is due to the unique attenuation characteristics of RF signals. Furthermore, when the BS and the SNs are not in the range of each other, the use of direct communication is not possible. However, to preserve the advantage of multi-hop communication, the routing algorithm should be energy-efficient and should not have high control messages complexity for constructing and maintaining the routing paths.

B. LOAD BALANCING

1) **Size of Clusters:** Random size of Clusters is not an energy-efficient method. Neither is the construction of equal clusters without paying attention to the distance from the CHs to the BS and the inter-cluster communication mode. Unequal clustering methods achieve more energy balancing by forming clusters with smaller sizes in the vicinity of the BS while using multi-hop communication. Minimizing the internal load (by assigning fewer member nodes to the CHs near the BS) helps these CHs saving energy for relaying the received data from farther CHs to the BS. When direct communication is used between the CHs and BS, unequal clustering approaches assign fewer member nodes to the CHs that are far from the SB. This is because those CHs may exhaust their energies very quickly compared to those close to the SB. This way, a fair distribution of energy dissipation between the different CHs in the network is assured.
2) **Network Re-clustering:** Dynamic clustering based on periodic network re-clustering is used to fairly balance the CH load between SNs. For this, the network operation is divided into rounds and at the beginning of each one, nodes exchange messages containing some information permitting to establish new clusters. Nevertheless, periodic re-clustering of the whole network is energy consuming and lessens the network lifetime, especially in centralized clustering approaches when SNs have to communicate with BS over long distances. Solutions to reduce the cost of network re-clustering are then needed. Some existing protocols suggest the election of two CHs per cluster. This is not sufficient

to equilibrate the load in the network since only few nodes will act as CHs. Other protocols propose the rotation of the role of CH within the cluster without changing its members. In this case, nodes with small amount of remaining energy or those located far from the BS may be elected as CHs. Likewise, nodes may be assigned to some distant CH instead of available closer ones, which will have a negative impact on the network lifetime. Energy estimation based clustering seems to be a good alternative to fix this problem provided that: i) an accurate energy estimation technique is used, and ii) all the energy consumption sources in a SN are considered. However, this technique can be used only in centralized or hybrid clustering, in which the BS or the CHs can collect energy information from the SN in some predefined rounds and predict the value of the nodes energy dissipation for the next rounds. The estimated values can be used to recluster the network without the need of exchanging messages.

3) **Round Duration:** A long round-time may cause the fast depletion of CHs energy, while a short one may result in a large waste of energy for frequent re-clustering and a non-optimal use of the CHs energy. The majority of existing protocols do not use thorough models or methods to dynamically set this duration but simply assign a static (sometimes arbitrary) time for all rounds independently of residual energy of CHs. Some use a simple model and set the round' durations such that all nodes act as CH once, and as non-CH in the other $(N/K-1)$ rounds (where N is the number of nodes in the network, K is the number of CHs). This approach is not effective for topologies where nodes are located in positions that are unsuitable for acting as CHs. Criteria for optimal rotation frequency is needed to reduce the network overhead of frequent CH rotation and avoid power draining of CHs in less frequent CH rotation.

C. PROTOCOL CHARACTERISTICS

1) **Protocol Complexity:** All clustering protocols have energy cost for the creation and the maintenance of the hierarchy. An energy-efficient clustering protocol should minimize both the number and the size of control messages (overhead).
2) **Data Aggregation:** The CHs may aggregate all the received data from their members and send only the meaningful information to the BS. In the case of multi-hop routing between the CHs, the intermediate CHs (relay CHs) can also perform further data aggregation. The clustering protocol should take benefit from this advantage for further optimizing the energy usage in the network.

V. REVIEW AND COMPARISON OF CLUSTERING PROTOCOLS

In the following, some canonical state-of-the-art WSN clustering protocols are reviewed. They are then compared according to the energy-oriented taxonomy proposed in Section IV.

A. REVIEW

The majority of the clustering protocols reviewed hereafter are based on heuristics and meta-heuristics. Heuristics are greedy problem-dependent algorithms in which some rules are proposed to elect the CHs and form clusters, using the intrinsic characteristics of the SNs. Meta-heuristics are problem-independent general framework algorithms, such as GA, SA, PSO, etc., that can be adapted to solve the clustering problem. Furthermore, according to the clustering methodology they use, clustering protocols are divided into three categories: distributed, centralized, and hybrid. In this subsection, representative protocols of each of these three categories are concisely described.

1) Distributed protocols

LEACH: LEACH by Heinzelman et al. [10] is one of the first clustering protocols proposed in the literature. It performs in rounds and uses a distributed probabilistic process to elect CHs in each round. Every SN picks up a random number between 0 and 1. If the number is less than a threshold $T(n)$, the SN announces itself as a CH for the current round. The threshold is based on a predetermined percentage of CHs in the network and the number of times the SN has been acting as a CH so far. Each non-CH SN chooses its CH according to the strength of signals it receives, i.e, it chooses the CH that can be reached using minimum energy transmission. In order to balance the load and the energy dissipation in the network, the role of CH is rotated periodically among the SNs by repeating the same process described above at the beginning of each round. Many recent solutions have been proposed based on LEACH. Some of these solutions are presented in the following.

LEACH-DT: A distance-based thresholds version of LEACH, LEACH-DT, has been proposed by Kang and Nguyen [41]. The key objective in LEACH-DT is to balance the energy dissipation in the network by considering the distance between the SN and the BS when calculating the probability of each SN for being CH. In LEACH-DT, the clusters are formed based on the distance between the SNs and the CHs, as in LEACH. The authors also proposed a multi-hop extension to LEACH-DT to avoid that remote CHs dissipate energy in direct communications with the BS.

LEACH-MF: In order to reduce packet loss while prolonging the network lifespan of mobile WSNs, Lee et al. proposed LEACH-MF [42]. This protocol modifies the CHs election algorithm of LEACH by using the fuzzy inference system with three inputs: SN remaining energy, moving speed, and pause time. SNs with higher remaining energy, slower moving speed, and longer pause time have a higher chance to get elected as CHs in the incoming round. By considering moving speed and pause time of each SN, stable connections for data dissemination are established, which can improve the packet delivery ratio. After electing the CHs, each SN joins its nearest CH as in LEACH.

LEACH-ERE: Lee et al. proposed another clustering solution based on LEACH, entitled LEACH-ERE [43]. In this protocol, an energy estimation model is used to calculate the expected residual energy of each SN after one round running as CH. The expected residual energy along with the remaining energy of the SN are then used by the fuzzy inference systems to compute the probability of a SN to become CH. SNs with higher probability values among their neighbors are elected as CHs for the incoming round. The other phases of LEACH-ERE are similar to those of LEACH.

ECPF: Taheri et al. [44] proposed the ECPF protocol that uses the fuzzy-logic [45], with SN degree and SN centrality as input. The SN degree is the number of neighbors of the SN divided by the network size. SN centrality is a value reflecting how central the SN is among its neighbors. In ECPF, each SN computes its cost via the fuzzy system. Then, each SN sets a delay timer inversely proportional to its remaining energy value. If the SN does not receive a CH announcement after the expiration of its timer, it broadcasts a tentative CH advertisement within its range. The SN having the least cost among the other tentative CHs becomes the final-CH. Every non-CH SN joins the final-CH with the smallest cost. The re-clustering is not performed at the beginning of each round, but only sporadically when a CH depletes a predefined fraction of its energy.

NARTC: Pal et al. proposed NARTC [46] that uses an approach to make adaptive round-time, which is calculated in function of the number of the alive nodes. When 50% of nodes in the network are alive, the round-time is calculated with the principle that all nodes should act as CH once, and as non-CH in the other $(N/K-1)$ rounds, where N is the size of the WSN and K is the number of CHs. Upon the death of 50% of nodes, the round-time becomes fixed. According to the authors, this is because the residual energy of SNs at that time will be very low, approximately $10-20\%$ from initial energy. Note that the technique of calculation of the dynamic round time proposed in NARTC is implemented with the LEACH protocol.

DSBCA: Contrary to the majority of existing clustering solutions that assume uniform distribution of the nodes in the network, DSBCA [47] focuses on WSN with stochastic distribution of SNs, for the purpose of load balancing. In DSBCA, the CHs election process is triggered by some randomly selected SNs that become temporary CHs. Each temporary CH calculates its cluster radius, k, as a function of its distance to the BS and the nodes distribution. It also calculates a weight which takes the SN remaining energy, density, and times of being elected as CH into account. The temporary CH with the highest weight value in its k-hop neighbors is then elected as final CH. Non-CH SNs send join-request to the nearest CHs (in terms of hop number). A threshold of cluster size is used by the CHs to decide whether or not the join request is accepted.

HL-FCM: Bouyer et al. have proposed a new hybrid protocol that uses LEACH and FCM algorithm [48]. This protocol tries to balance the intra-cluster energy to maximize the network lifespan and the network coverage by tackling two problems of LEACH: i) CHs located far from the BS dissipate more energy in data transmission, and ii) The random selection of CHs that leads to an unbalanced network if a node of low energy is elected as CH. The FCM algorithm is used to change the LEACH protocol parameters for the optimal values, as well as the locations of the CHs. Each SN is then assigned to its nearest CH.

HT2HL: HT2HL by Alharthi et al. [49] enhances the performance of LEACH in heterogeneous WSN and deals with some of its drawbacks. It proposes techniques to i) permit the direct communication of the SNs near the BS, i.e. those SNs will send their data to the BS without passing by a CH, ii) enhance the CHs election thresholds to take into account the location of the SN and its energy-efficiency, iii) thresholds are incorporated to reduce the number of transmissions. Note that the cluster formation is similar to that of LEACH.

ECFU: Radhika et al. proposed ECFU [50] that attempts to improve the lifespan by using a fuzzy-logic and machine learning based on data reduction. The SN can take up the role of CH, ancillary node, or cluster member. The nodes with the largest energy are picked as CHS and each SN join the nearest CH or ancillary node. The role of the ancillary is to manage the cluster once the energy of the CH drains out. To reduce the overhead of network re-clustering, the authors propose the use of the fuzzy inference system that allows deciding when a re-clustering will take place. Moreover, ECFU employs a machine learning based data transmission in which the SNs send only dissimilar data to the CHs. Consequently, the number of data transmission is decreased which permits further energy saving.

MOFCA: In order to address energy hole and hot-spot problems, Sert et al. proposed MOFCA [51]. It is a distribution-agnostic protocol and considers both stationary and evolving WSNs. MOFCA performs in rounds. At the beginning of each one, a random number is generated by each SN and compared with a threshold to decide the provisional CHs. The fuzzy inference engine is then utilized to calculate the radius over which provisional CHs will compete using the following parameters: SN remaining energy, SN density and d(SN, BS). SN density is calculated and broadcasted by the BS at the beginning of each round. After the election of the final CHs, basing on the remaining energy of the candidate CHs, each SN joins its nearest CH. Due to the method used to calculate the competition radius in MOFCA, the WSN is divided into unequal clusters.

TTDFP: Sert et al. proposed TTDFP [52] as another solution to deal with the hot-spot problems and optimizing data aggregation. This protocol considers energy efficiency in both clustering and routing phases. The CHs election and cluster formation processes are almost similar to that of MOFCA [51], except for what follows. TTDFP uses the SN connectivity parameter instead of SN density. SN

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3092509, IEEE Access

connectivity is calculated by the SNs, which makes this protocol fully-distributed. To further optimizing the WSN performance metrics, TTDFP uses the SA algorithm to tune the fuzzy clustering parameters. FL is also used by TTDFP to establish the routing paths connecting the CHs to the BS with the objective of reducing and balancing the overall energy consumed in the WSN.

EEFL-CH: In order to optimize the energy consumption in the WSN, El Alami et al. make use of the fuzzy inference system to enhance the CHs election phase of LEACH by proposing of EEFL-CH [53]. In this protocol, three fuzzy parameters are considered to calculate the chance of a SN to become CH in each round. The parameters are, i) SN residual energy, ii) its distance to the BS and iii) the expected efficiency. The latter is defined as a ratio between the expected residual energy of a SN and the expected average energy of the cluster. The expected residual energy of the SNs is estimated using an energy consumption model.Cluster formation phase and data dissemination are similar to that of LEACH.

EAUCA: EAUCA [54], by Chauhan et al., investigates the energy hole problem and aims to extend the network lifespan by creating unequal sized clusters. The unequal clustering is achieved by calculating the competition radius of the candidate CHs based on the SN remaining energy and its distance to the BS. The final CHs are then selected by considering both residual energy and degree of SNs and each SN joins the CH requiring minimum transmission energy. EAUCA proposes a multi-hop routing algorithm between the CHs and the BS to further energy preservation. Relay nodes are chosen in accordance with the SN remaining energy, degree, and distance to BS.

HDDS: Mazumdar et al. proposed HDDS [55], an energy-efficient hierarchical protocol aiming at enhancing the network coverage and supporting the dynamic changes in the WSN topology. To select the optimal set of CHs, every SN triggers a delay timer which is tuned to match the better candidates for the CH role with the smaller values. These waiting values are computed based on the SN residual energy, coverage, and distance to the BS. A SN elects itself CH once its timer expires or if it does not receive any advertisement message from an elected CH. To deal with the hot spot problem, this protocol calculates the cluster range of an elected CH in such a way that CHs closer to the BS have smaller cluster radius. A multi-objective optimization function is used by the non-CH SNs to decide which CH to join. This function considers the CH residual energy and the transmission energy that will be spent by the SN to reach this CH. A multi-hop routing is also proposed in this protocol with the aim of optimizing energy usage and enhancing the reliability. Moreover, re-clustering is not performed in each round in HDDS, but at regular time intervals.

UCMR-FL: In order to minimize and balance the total energy dissipated in the network, Adnan et al. introduced UCMR-FL [56]. This protocol proposes the use of the SN remaining energy, its distance to the BS and its concentration (number of adjacent nodes) as inputs of a FL-based algorithm. The outputs of this algorithm are both the SN chance for being elected CH and its competition radius. The algorithm enables every non-CH SN to join the nearest CH. By calculating the CH competition radius, UCMR-FL forms unequal clusters in the network to avoid the hot spot problem. In UCMR-FL, CHs relay the collected data to the BS through a multi-hop routing algorithm.

ECH: To achieve an efficient energy usage, El Alami et al. proposed ECH [57] in which the issue of data redundancy is tackled by using a sleeping/waking mechanism for neighboring and overlapping SNs. At the beginning of each round of the EHC protocol, groups of neighboring SNs having overlapping sensing ranges are formed. A subset of SNs is then chosen in each group to be awake while the other SNs enter into sleepy mode. SNs do not have neighbors in their sensing range will remain awake till they exhaust their energy. After that, some of the waking SNs elect themselves randomly as candidate CHs and the optimal number of clusters is calculated using an energy prediction mechanism. If the number of candidate CHs in a group is strictly smaller than the desired number of CHs, then re-election of candidate CHs is performed based on the average distances between the SNs and all the candidate CHs. Otherwise, a minimization of the number of candidate CHs is carried out considering both energy and distance parameters and using the heuristic fuzzy method. To form clusters, all the elected CHs along with the BS broadcast a packet in the network. SNs use the signal strength when receiving these packets and join the transmitter engendering the maximum value.

2) Centralized Protocols

LEACH-C: LEACH-C [11], by Heinzelman et al., is a centralized version of LEACH that has been proposed to deal with the random selection and non-uniform distribution of CHs. At the beginning of each round, each SN communicates its remaining-energy and position value to the BS. Thereby, the BS calculates the average energy of the network. SNs having residual energy higher than the average are eligible to compete for CH role in the next round. A SA algorithm is then applied by the BS on the eligible set of nodes to determine the CHs for the incoming round. This algorithm attempts to reduce the total energy that will be dissipated by the non-CH SNs to transmit their data to their CHs. Each non-CH SN will be member of the nearest CH' cluster. Finally, the list of resulted CHs and their respective clusters are broadcasted by the BS.

LEACH-F: LEACH-F [11], by Heinzelman et al., is based on LEACH-C (the same CHs election and cluster formation processes). In LEACH-F, the clusters are formed at the network initialization and become fixed throughout the whole network lifespan. The BS creates and broadcasts a schedule between the members of each cluster to permit the intra-cluster rotation of the CH role. This is to reduce the cost of re-clustering the network at each round.

LEACH-CE: LEACH-CE [58], by Kim et al., has also been proposed to reduce the cost of network re-clustering. It proposes an energy prediction technique by defining a set-up phase that takes place every eight rounds. In this phase, nodes send their energy information at the beginning of two successive rounds. Based on this information, the BS can calculate the average of the energy consumed by CHs and member nodes in one round, and thus their residual energy. This energy estimation model is used in the upcoming eight rounds to avoid exchanging energy levels. Note that LEACH-CE adopts the CHs election and cluster formation phases proposed in LEACH-C.

D-LEACH-F: The objective of D-LEACH-F [59] by Azim et al. was to mitigate the fixed round time problem of LEACH-F and avoid fast CHs' batteries depletion. They proposed a method to model the rounds' durations and allow all nodes to act as CH once, and as non-CH in the other $(N/K-1)$ rounds, where N is the number of nodes in the network and K is the number of CHs. The general operations of this protocol are similar to LEACH-F, except that after broadcasting the clustering scheme, the BS waits until getting current energy information from CHs. It then calculates and broadcasts the round duration with the CH identifier to the corresponding cluster members.

HMM-PSO: Goudarzi et al. proposed HMM-PSO protocol [60] that combines HMM and PSO algorithms. This protocol also performs in rounds. In the set-up phase of the first three rounds, all nodes send information about their remaining energy level and location to the BS. Based on this information, the BS estimates the energy consumed by each node and its residual energy using the HMM, which has been optimized by the PSO [61] to increase the accuracy of prediction. PSO is also used to select the suitable set of CHs that can minimize the ICD, and to optimize energy usage. After the election of the CHs, every non-CH SN is assigned to its nearest CH.

EBUC: Jiang et al. proposed EBUC [62] to deal with the hot-spot problem of multi-hop WSN and for balancing the energy consumption in the network. The network is partitioned into clusters of unequal sizes, such that clusters closer to the BS have smaller sizes. In this manner, the CHs of those clusters can preserve energy for assuring the routing task. EBUC proposes an algorithm based on PSO to create optimized unequal clusters. This algorithm considers three criteria to elect the CHs namely, energy, ICD and distance between the SN and the BS. Moreover, two criteria are used in the cluster formation phase, which are the distance between the SN and the CH and the distance between the CH and the BS. EBUC also adopts an energy-aware multi-hop routing algorithm between the CHs and the BS to reduce the energy dissipated by the CHs. Further, EBUC uses an energy prediction method to estimate the SNs energy level at the beginning of each round and thus alleviate the cost of frequent network re-clustering.

IEEHCS: Patra et al. have proposed IEEHCS [63]. In the set-up phase of the first round of this protocol, the BS calculates the optimal number of CHs as a function of some parameters such as the network size and the transmission range of each SN. It then elects the appropriate CHs according to the remaining energy of each SN, number of neighbors within its transmission range and minimum distance between two CHs. After that, the BS form the clusters basing on the distance separating the SNs and the elected CHs. Re-clustering is not performed automatically at every round in IEEHCS. Instead, the residual energies of CHs are used to decide whether to re-cluster the network or to use the existing CHs. If the remaining energy of every CH does not fall below the threshold (set by BS according to the average remaining energy of the network), then no re-clustering is triggered and the same CHs continue for the next round. Otherwise, the role of this CH is shifted to the nearest member node whose remaining energy is above the threshold. Finally, when the remaining energy levels of all CHs fall below the energy threshold, the whole WSN is re-clustered.

HSACP: Hoang, et al. have proposed a real-time implementation of the HSACP protocol [64]. This is one of the very few clustering protocols that have been implemented on a WSN test-bed. In this implementation, the SNs are deployed in an indoor environment to monitor the ambient temperature for fire detection. The clustering algorithm is executed by the BS which consists of a computer connected to a gateway node. The clustering algorithm uses HSA to optimize the cluster formation and the CH selection. SNs with the maximum of residual energy and that minimize the ICD are elected as CHs, and the other SNs join the CH requiring the minimum communication energy.

TPSO-CR: In TPSO-CR [40], Elhabyan et al. introduced two linear programming models to the clustering and routing problems. Two algorithms based on the particle swarm optimization meta-heuristic have been proposed to solve the models. Energy consumption, cluster quality, and network coverage are the criteria considered by TPSO-CR in the CHs election process. Distance between the SNs and the CHs is considered when forming the clusters. An energy-efficient fitness function is also proposed to find the optimal routing tree connecting CHs to the BS.

ETPSO-CR: ETPSO-CR [16] is an enhanced version of TPSO-CR in which Merabtine et al. proposed three techniques to prolong the network lifetime. The first one introduces a factor in the clustering objective function to balance the geographical distribution of the CHs, for the purpose of balancing the energy consumed by the CHs and minimizing the total energy dissipated by non-CH-nodes. The second one adapts the round-time proportionally to the CHs residual energy, to achieve an optimal use of the nodes batteries. The third technique consists of rotating the role of CH amongst the network based on an energy prediction model, which allows to save considerable energy that would be spent in exchanging energy states information.

CECP: Hatamian et al. proposed CECP [65] that uses a GA to select the optimal set of CHs. SNs to be elected as CHs are those having remaining energy higher than the network average energy, that are not outlier nodes, and that have the maximum sum of edges' weights in the graph. The latter is constructed between the SNs which are in the radio communication range of each other. A weight is assigned to each edge of this graph, and then the sum of weights for edges connected to each SN is calculated.

OSC: Merabtine et al. focused on periodic traffic and proposed OSC [13], which is a centralized protocol that uses a one-step off-line cluster computation algorithm. In this algorithm, all the clustering schemes and their respective durations are calculated by the BS once at the network initialization. This provides the BS a global vision and enables it to reach better clustering schemes with adapted rounds' durations. OSC makes use of an energy prediction mechanism to alleviate the cost of periodic online re-clustering. Further, OSC introduces a new weight function to evaluate the chance of SNs to become CHs in a round. The aim of this function is to select the set of CHs that minimizes energy consumption of all the SNs in the network. Then, each non-CH SN joins the first elected CH among its neighbors to form the clusters.

LPOBC: Tabatabaei et al. used LPO algorithm and FL to introduce LPOBC [66]. In this protocol, the BS selects the best set of CHs in the network according to two parameters; the residual energy of the nodes and their distance from the BS. The BS informs the nodes elected as CHs, which broadcast advertisement messages within their radio range. The other SNs in the network join the closest CH. After the formation of clusters, the BS constructs a DVB of CHs to permit an efficient data routing.

3) Hybrid Protocols

TTCH-WSN: Meenakshi et al. proposed a clustering protocol that investigates the idea of electing two CHs in each cluster [67]. The first CH receives the collected data from its cluster members and transmits it to the second CH which is in charge of data aggregation and communication with the BS. The authors argued that this way the load of CHs will be reduced. Besides, TTCH-WSN proposes a whole network re-clustering every 100 rounds. The election of the first CH and the creation of the clusters are carried out by the BS, while the selection of the second CH is done by the first CH through a distributed algorithm. The BS takes into account the residual energy of each SN, whether it has been elected as CH in the last $1/P$ rounds or not and the total desired number of CHs to select the set of SNs to be elected as CHs for the incoming round. It then assigns every SN to its nearby CH to form clusters.

MMDCP: Abdel-Hady et al. considered the end-to-end-delay and proposed MMDCP [68]. This protocol aims to extend the lifetime of the WSN and minimize the end-to-end delay through optimal selection of CHs. It is a hybrid clustering protocol in which some nodes called (coordinators) are deployed in different zones of the WSN. Each coordinator organizes its zone into different clusters. MMDCP assigns the number of the lower level CHs and the leaf SNs in the network so as to minimize the end-to-end delay. It chooses the CHs by calculating the sum of the minimum distances of every SN to all other SNs in the network and sorting them in ascending order. The first 5% of the number of SNs in the network are selected from the top of the list, i.e., the SNs with minimum distances to all other SNs in the network. MMDCP introduces a novel balancing algorithm for assigning each non CH SN to their CH. This algorithm takes into account the residual energy of CHs and the distance between the SN and CHs when assigning SNs to their CHs. Note that both CH selection and member partitioning are executed by the coordinators.

MS-routing-Gi: El Alami et al. have considered WSNs with mobile sink and proposed the MS-routing-Gi protocol [69]. The first step of MS-routing-Gi is carried out by the BS and aims to divide the WSN into grids (inner vs. clustered grids). SNs belonging to an inner grid communicate directly with the sink, while CHs are elected to assure communication between SNs and the sink in the clustered grids. The aim of such architecture is to achieve balanced energy consumption in the network. To elect CHs and form clusters, SNs of each grid cooperate a distributed manner using a probabilistic algorithm similar to that proposed in LEACH. In each round of the MS-routing-Gi operations, the sink changes its position according to a predetermined trajectory. The latter is calculated based on a cost function defined as the ratio between the energy level of each CH and the size of its cluster. The authors also tackle the issue of packet loss by formulating it as an optimization problem with an objective function aiming at minimizing the total packet loss rate in the network.

B. DISCUSSION

This subsection discusses the reviewed clustering protocols using the proposed energy-oriented taxonomy (Figure 4). Tables 2, 3 and 4 highlight how each protocol considers the factors identified in the aforementioned taxonomy. From these tables it can be noticed that energy efficiency is the predominating criterion used for selecting the CHs in all clustering methodology (distributed, centralized, or hybrid). This is reflected by considering the remaining energy of the SNs in most existing protocols, such as LEACH-C [11], TPSO-CR [40] and ETPSO-CR [16]. In OSC [13], both the remaining energy of the candidate SN and the average energy required to transmit a packet to that SN from its neighbors are used. In addition to the residual energy, LEACH-ERE [43] evaluates the energy efficiency of a SN by estimating its expected remaining energy after acting as CH for one round. In EEFL-CH [53], a ratio between the expected residual energy of each SN and the expected average energy of the cluster is calculated in order to select the energy efficient set of CHs. There are other criteria for CHs election that indirectly reflect the energy efficiency. For instance, LEACH-DT [41], the hybrid-protocol in [48], HT2HL [49], EBUC [62], OSC [13], LPOBC [66], MMDCP [68] and MS-routing-Gi [69] use the $d(SN, BS)$ which highly impacts the energy to be dissipated if the SN is chosen as CH. Further, the ND is used by clustering protocols that consider data aggregation, such as OSC [13], IEEHCS [63], ECPF [44]. This is to push SN with a higher ND for the CH role so that it can collect high number of data packets and then aggregate them into single packet. This obviously minimizes the size of data transmitted to the BS and reduces the amount of the required transmission energy. Moreover, HMM-PSO [60], EBUC [62] and HSACP [64] use the ICD criterion to minimize the transmission energy of the non CHs SNs. To balance the energy consumption between the SNs and prevent their earlier death, DSBCA [47], HT2HL [49], TTCH-WSN [67] consider the criterion CH times. The majority of the reviewed protocols combine more than one criterion for the CHs election, such as ECH [57] and CECP [65] which consider both the remaining energy of the SNs and the distances between the SNs. These protocols utilize the same criteria but in completely different manners and while targeting one objective; minimizing the overall dissipated energy. DSBCA [47], EEFL-CH [53], UCMR-FL [56], ECH [57], OSC [13], ECPF [44], HT2HL [49], IEEHCS [63], TTCH-WSN [67] and EBUC [62] make use of the largest number of energy related criteria in the CHs election process.

Certainly, the choice of the criteria for the CHs election process is crucial. Nevertheless, the algorithm used to find these CHs has also an important repercussion on the energy effectiveness of the chosen CHs. Centralized clustering protocols take benefit of the complete topoligical information and the strong computation ability of the BS and use meta-heuristic algorithms such as GA in CECP [65], SA in LEACH-C [11], PSO in HMM-PSO [60], EBUC [62], TPSO-CR [40], ETPSO-CR [16], etc. Fuzzy logic is used in the distributed protocols to deal with uncertainty among multiple parameters used in the CHs election, e.g., in MOFCA [51], TTDFP [52], UCMR-FL [56].

For the number of CHs factor, the majority of clustering protocols fix it to 5% or 10% of the network size. Only the hybrid-protocol in [48], IEEHCS [63] and ECH [57] propose techniques to calculate the optimal number of CHs. In EAUCA [54], the number of CHs is not explicitly calculated but it depends on the input parameters, so it is adaptive. ECPF [44], DSBCA [47], ECFU [50], OSC [13], MOFCA [51], HDDS [55] and TTDFP [52] do not take in consideration this factor. Instead, they elect a random number of CHs in each round. To minimize the energy dissipation of the non-CH SNs, $d(SN, CH)$ is used in the majority of the reviewed protocols to decide how the SNs join CHs. In addition to the $d(SN, CH)$, ECH [57] considers $d(SN, BS)$ in order to allow the SNs to communicate directly with the BS if their distance from the BS is smaller than the one from the nearby elected CHs, which can preserve the SNs energy. DSBCA [47], EBUC [62] and MMDCP [68] add other criteria related to network density and the desired size of clusters to balance the energy usage between the CHs. In HDDS [55], a multi-objective function which takes into account the CHs remaining energy and the transmission energy that will be spent by a SN to reach a CH is used to decide which CH to join. This has the advantage of optimizing the energy dissipation of both CH and non-CH SNs.

Furthermore, all the studied protocols (except for DSBCA [47]) adopt the single-hop communication paradigm between the SNs and their CHs. Likewise, single-hop is also used in several protocols for inter-cluster communication. However, some clustering protocols add multi-hop routing between CHs and BS, including LEACH-DT [41], ECPF [44], EBUC [62], TPSO-CR [40], ETPSO-CR [16], LPOBC [66], TTCH-WSN [67], MMDCP [68], MOFCA [51], TTDFP [52], EAUCA [54], HDDS [55], UCMR-FL [56]. Since the transmission energy is proportional to the transmission distance, the use of multi-hop routing between the CHs and the BS can improve the energy efficiency of the WSN. DSBCA [47], hybrid-protocol [48], EBUC [62], MMDCP [68], MOFCA [51], EAUCA [54], HDDS [55], TTDFP [52] and UCMR-FL [56] give importance to the cluster size by creating unequal clusters to preserve the CHs energy, balancing the load between the CHs and avoiding the energy hole problem. In DSBCA [47], MOFCA [51], EAUCA [54], UCMR-FL [56] and HDDS [55], the unequal clustering is achieved by calculating the CHs competition radius relying on some criteria such as the energy level of the SN, the ND and d(SN, BS).

Many of the surveyed clustering protocols attempt to alleviate the cost of network re-clustering. LEACH-F [11] and DSBCA [47] make use of local re-clustering. TTCH-WSN [67] elects two CHs in every cluster. Other protocols propose the use of energy prediction mechanisms such as LEACH-CE [58], D-LEACH-F [59], HMM-PSO [60], EBUC [62], ETPSO-CR [16], OSC [13]. Moreover, some protocols consider sporadic network re-clustering to grant the predicted models with accuracy, e.g., TTCH-WSN [67] IEEHCS [63], LEACH-CE [58], ECFU [50], ECPF [44].

The round duration is modeled and calculated in only four protocols reviewed here, which are NARTC [46], D-LEACH-F [59], ETPSO-CR [16] and OSC [13]. All the remaining protocols use a fixed round duration. This is not optimal, and optimal setting of the round duration is an important factor to achieve energy efficiency. ECPF [44], DSBCA [47], ECFU [50], LEACH-CE [58], HMM-PSO [60], EBUC [62], IEEHCS [63], ETPSO-CR [16], OSC [13], are the protocols that have the lowest overhead complexity. This is because they adopt some techniques for mitigating the cost of the periodic network re-clustering. Finally, it is remarkable that almost all of the surveyed clustering protocols take advantage of data aggregation for further energy optimization, except LPOBC [66], DSBCA [47], EBUC [62].

VI. CONCLUSION AND PERSPECTIVES

This work has been elaborated as an effort to facilitate the study of clustering protocols proposed for WSNs. A comprehensive review of the state-of-the art approaches has been provided in this paper, and a global feature-based clustering classification has been proposed. The factors that impact the energy-efficiency of the clustering protocols have been identified and emphasized in a novel energy-oriented taxonomy. This can serve as a tool to theoretically evaluate existing clustering protocols, as well as a guideline to propose more energy-efficient clustering solutions. Some existing clustering protocols have been reviewed and compared in line with the proposed taxonomy. We noticed that some algorithms such as those based on meta-heuristics and FL are useful for centralized protocols to create optimized clusters in terms of number and size. They also help electing the optimal set of CHs with the nearest nodes as members. This way, member nodes communicate directly with their CHs while a simple energy aware routing mechanism can be used to route data from CHs to BS. The cost of creating such centralized solution is summed up in the information messages sent from the nodes to the BS at the network initialization and the message broadcasted by the BS to announce the formed clusters. The cost of network re-clustering can be mitigated through the use of an accurate and realistic energy prediction model that takes into consideration all energy consumption sources, such as idle listening, overhearing, collisions, etc. Further, re-clustering the whole network can be triggered sporadically to cover the nodes failure or the erratically energy exhaustion. The time when a re-clustering should take place may be periodic, or calculated by the fuzzy inference system. The round duration should be adjusted in such a way to optimize the use of batteries and to avoid premature power drainage of CHs. Data aggregation should be explored for further energy conservation, as well as cross-layer approaches that incorporate the MAC protocol. For instance, CHs can establish a schedule for their member nodes using TDMA, and then each member node can switch off its radio once its data is sent. Machine learning techniques can also be used to classify data based on their similarity in order to lessen the number of messages being transmitted from the member nodes to CHs. As a perspective to this is work, we plan to conduct a comparative study on the performance of meta-heuristic algorithms (GA, PSO, SA, bats algorithms, etc.), as well as fuzzy-logic methods, when applied to the clustering problem. This will enable to identify the most appropriate method for energy efficient clustering.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3092509, IEEE Access

Protocol	Protocol design					Load balancing					Protocol characteristics	
	CH selection		Member affectation	Communication pattern		Clusters size	Network re-clustering			Round duration	Complexity	Aggregation
	Criteria	Number		SN to CH	CH to BS		Dynamic	Several CHs	Estimation			
LEACH [10]	random	5% network size	d(SN, CH)	single-hop	single-hop	random	yes	no	no	fixed	low	yes
LEACH-DT [41]	AB (d(SN, BS))	5% network size	d(SN, CH)	single-hop	multi-hop	random	yes	no	no	fixed	low	yes
LEACH-MF [42]	AB(E, moving speed, pause time)	5% network size	d(SN, CH)	single-hop	single-hop	random	yes	no	no	fixed	low	yes
LEACH-ERE [43]	AB(E, expected residual energy)	5% network size	d(SN, CH)	single-hop	single-hop	random	yes	no	no	fixed	low	yes
ECPF [44]	AB (E, fuzzy cost with ND and centrality)	random	d(SN, CH)	single-hop	multi-hop	random	yes	no	no	fixed	very low	yes
NARTC [46]	random	5% network size	d(SN, CH)	single-hop	single-hop	random	yes	no	no	calculated	low	yes
DSBCA [47]	AB (E, density, CH times)	random	density, d(SN, CH)	multi-hop	single-hop	unequal	no	no	no	fixed	very low	no
HL-FCM [48]	AB (E, d(SN, BS))	calculated (FCM)	d (SN, CH)	single-hop	single-hop	unequal	yes	no	no	fixed	low	yes
HT2HL [49]	AB (d(SN, BS), E, CH times)	5% network size	d (SN, CH)	single-hop	single-hop	random	yes	no	no	fixed	low	yes
ECFU [50]	AB (E)	random	d(SN, CH) or ancillary	single-hop	single-hop	random	yes	no	no	fixed	very low	yes
MOFCA [51]	AB(E, density, d(SN, BS)) for competition radius calculation and E for the final CHs	random	d(SN, CH)	single-hop	multi-hop	unequal	yes	no	no	fixed	low	yes
TTDFP [52]	AB(E, connectivity, d(SN, BS)) for competition radius calculation and E for the final CHs	random	d(SN, CH)	single-hop	multi-hop	unequal	yes	no	no	fixed	low	yes
EEFL-CH [53]	AB(E, expected efficiency , d(SN, BS))	5% network size	d(SN, CH)	single-hop	single-hop	random	yes	no	no	fixed	low	yes
EAUCA [54]	AB(E, d(SN, BS)) for competition radius calculation and AB(E, ND) for CHs	variable	d(SN, CH)	single-hop	multi-hop	unequal	yes	no	no	fixed	low	yes
HDDS [55]	AB(E, coverage, d(SN, BS))	random	AB(E, d(SN, CH))	single-hop	multi-hop	unequal	yes	no	no	fixed	low	yes
UCMR-FL [56]	AB(E, SN concentration, d(SN, BS))	5% network size	d(SN, CH)	single-hop	multi-hop	unequal	yes	no	no	fixed	low	yes
ECH [57]	random for candidate CHs, AB(Energy-efficiency, d(SN, candidate CH), d(SN, BS))	calculated	d(SN, CH) and d(SN, BS)	single-hop	single-hop	random	yes	no	no	fixed	low	yes

TABLE 2: Comparison of some distributed clustering protocols in WSNs from an energy-efficiency point of view.

Protocol	Protocol design					Load balancing					Protocol characteristics	
	CH selection		Member affectation	Communication pattern		Clusters size	Network re-clustering			Round duration	Complexity	Aggregation
	Criteria	Number		SN to CH	CH to BS		Dynamic	Several CHs	Estimation			
LEACH-C [11]	AB (E)	5% network size	d(SN, CH)	single-hop	single-hop	random	yes	no	no	fixed	high	yes
LEACH-F [11]	AB (E)	5% network size	d(SN, CH)	single-hop	single-hop	random	no	schedule	no	fixed	very low	yes
LEACH-CE [58]	AB (E)	5% network size	d(SN, CH)	single-hop	single-hop	random	yes	no	yes	fixed	very low	yes
D-LEACH-F [59]	AB (E)	5% network size	d(SN, CH)	single-hop	single-hop	random	no	no	yes	calculated	low	yes
HMM-PSO [60]	AB (E, ICD)	predefined k	d(SN, CH)	single-hop	single-hop	random	no	no	yes	fixed	very low	yes
EBUC [62]	AB (E, ICD, d(SN, BS))	5% network size	d(SN, CH), d(CH, BS)	single-hop	multi-hop	unequal	no	no	yes	fixed	very low	no
IEEHCS [63]	AB (E, ND, distance between CHs)	calculated	d(SN, CH)	single-hop	single-hop	random	yes	no	no	fixed	very low	yes
HSACP [64]	AB (E, ICD)	predefined k	d(SN, CH)	single-hop	single-hop	random	yes	no	no	fixed	high	yes
TPSO-CR [40]	AB (E, cluster quality, network coverage)	5% network size	d(SN, CH)	single-hop	multi-hop	random	yes	no	no	fixed	high	yes
ETPSO-CR [16]	AB (E, cluster quality, network coverage, distance between CHs)	5% network size	d(SN, CH)	single-hop	multi-hop	random	no	no	yes	calculated	very low	yes
OSC [13]	AB (E, ND, d(SN, BS), the average energy required to transmit a packet to that SN from its neighbors)	random	the node joins the first elected CH among its neighbors	single-hop	single-hop	random	no	no	yes	calculated	very low	yes
LPOBC [66]	AB (E, d(SN, BS)	10% network size	d(SN, CH)	single-hop	multi-hop	random	no	no	no	fixed	high	no
CECP [65]	AB (E, not outlier nodes, sum of edges weights in the graph)	5% network size	not specified	single-hop	single-hop	random	yes	no	no	fixed	high	yes

TABLE 3: Comparison of some centralized clustering protocols in WSNs from an energy-efficiency point of view.

Protocol	Protocol design					Load balancing					Protocol characteristics	
	CH selection		Member affectation	Communication pattern		Clusters size	Network re-clustering			Round duration	Complexity	Aggregation
	Criteria	Number		SN to CH	CH to BS		Dynamic	Several CHs	Estimation			
TTCH-WSN [67]	AB (E, no repetition of being CH for the last 1/P rounds, total number of CHs)	10% network size for the two types of CHs	d(SN, CH)	single-hop	multi-hop	random	yes	yes	no	fixed	low	yes
MMDCP [68]	AB (d(SN, BS))	5% network size	E and d(SN, CH)	single-hop	multi-hop	unequal	yes	no	no	fixed	low	yes
MS-routing-Gi [69]	AB (d(SN, BS)) for grids and random for CHs	5% network size	d(SN, CH)	single-hop	single-hop	equal	yes	no	no	fixed	low	yes

TABLE 4: Comparison of some hybrid clustering protocols in WSNs from an energy-efficiency point of view.

REFERENCES

[1] L. K. Alazzawi, A. M. Elkateeb, A. Ramesh, and W. Aljuhar. Scalability analysis for wireless sensor networks routing protocols. In 22nd International Conference on Advanced Information Networking and Applications, pages 139–144, 2008.

[2] Bhushan Trivedi Nilesh Modi, Pramode Verma. advances in intelligent systems and computing. In Proceedings of International Conference on Communication and Networks comNet 2016. Springer, 2016.

[3] B. Sundararaman, U. Buy, and A. D Kshemkalyani. Clock synchronization for wireless sensor networks: a survey. Ad hoc Networks, 3(3):281–323, 2005.

[4] Robert M. Curry and J. Cole Smith. A survey of optimization algorithms for wireless sensor network lifetime maximization. Computers and Industrial Engineering, 101:145 – 166, 2016.

[5] V Krishnaveni M Ezhilarasi. A survey on energy and lifetime in wireless sensor networks. Computer Reviews Journal, 1(1):30–51, 2018.

[6] Ameer Ahmed Abbasi and Mohamed Younis. A survey on clustering algorithms for wireless sensor networks. Computer Communications, 30(14–15):2826 – 2841, 2007.

[7] Geetha. V, Pranesh.V. Kallapur, and Sushma Tellajeera. Clustering in wireless sensor networks: Performance comparison of leach & leach-c protocols using ns2. Procedia Technology, 4:163 – 170, 2012. 2nd International Conference on Computer, Communication, Control and Information Technology(C3IT-2012) on February 25 - 26, 2012.

[8] M. Omari and S. Laroui. Simulation, comparison and analysis of wireless sensor networks protocols: Leach, leach-c, leach-1r, and heed. In 2015 4th International Conference on Electrical Engineering (ICEE), pages 1–5, 2015.

[9] M. Bakni, L. Manuel, M. Chacon, Y. Cardinale, and G. Terrasson et al. Wsn simulators evaluation: An approach focusing on energy awareness. International journal of wireless & mobile networks (IJWMN), 11(6):1–20, 2020.

[10] Anantha Chandrakasan Wendi Rabiner Heinzelman and Hari Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. In Proceedings of the 33rd Hawaii International Conference on System Sciences, 2000.

[11] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan. An application-specific protocol architecture for wireless microsensor networks. IEEE Transactions on Wireless Communications, 1(4):660–670, 2002.

[12] Majid Hatamian, Hamid Barati, Ali Movaghar, and Alireza Naghizadeh. Cgc: Centralized genetic-based clustering protocol for wireless sensor networks using onion approach. Telecommun. Syst., 62(4):657–674, 2016.

[13] Nassima Merabtine, Djamel Djenouri, Djamel Eddine Zegour, Elseddik Lamini, Rima Bellal, Imene Ghaoui, and Nabila Dahlal. One-step clustering protocol for periodic traffic wireless sensor networks. In 26th Wireless and Optical Communication Conference, (WOCC), pages 1–6, 2017.

[14] Lina Xu, Rem W. Collier, and Gregory M. P. O'Hare. A survey of clustering techniques in wsns and consideration of the challenges of applying such to 5g iot scenarios. IEEE Internet of Things Journal, 4(5):1229–1249, 2017.

[15] Sunil Kumar Singh, Prabhat Kumar, and Jyoti Prakash Singh. A survey on successors of LEACH protocol. IEEE Access, 5:4298–4328, 2017.

[16] Nassima Merabtine, Djamel Djenouri, Djamel Eddine Zegour, Billel Boumessaidia, and Abdelghani Boutahraoui. Balanced clustering approach with energy prediction and round-time adaptation in wireless sensor networks. International Journal of Communication Networks and Distributed Systems (IJCNDS)., 22(3), 2019.

[17] B. P. Deosarkar, N. S. Yadav, and R. P. Yadav. Clusterhead selection in clustering algorithms for wireless sensor networks: A survey. In International Conference on Computing, Communication and Networking, pages 1–8, 2008.

[18] P. Kumarawadu, D. J. Dechene, M. Luccini, and A. Sauer. Algorithms for node clustering in wireless sensor networks: A survey. In 4th International Conference on Information and Automation for Sustainability, pages 295–300, 2008.

[19] C. Jiang, D. Yuan, and Y. Zhao. Towards clustering algorithms in wireless sensor networks-a survey. In IEEE Wireless Communications and Networking Conference, pages 1–6, 2009.

[20] Sanjeev Kumar Gupta, Neeraj Jain, and Poonam Sinha. Article: Clustering protocols in wireless sensor networks: A survey. International Journal of Applied Information Systems, 5(2):41–50, 2013.

[21] O. Boyinbode, H. Le, A. Mbogho, M. Takizawa, and R. Poliah. A survey on clustering algorithms for wireless sensor networks. In 13th International Conference on Network-Based Information Systems, pages 358–364, 2010.

[22] Xuxun Liu. A survey on clustering routing protocols in wireless sensor networks. Sensors, 12:11113–11153, 2012.

[23] M. Mehdi Afsar and Mohammad-H. Tayarani-N. Clustering in sensor networks: A literature survey. Journal of Network and Computer Applications, 46:198 – 226, 2014.

[24] Sumedha Sirsikar and Kalyani Wankhede. Comparison of clustering algorithms to design new clustering approach. Procedia Computer Science, 49:147 – 154, 2015.

[25] P. Pandey and P. Srivastava. Efficient algorithm for cluster head selection in wireless sensor networks: A comparative study. In 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pages 2860–2864, 2016.

[26] N. Kaur, D. K. Grewal, and S. K. Shankar. Typical and atypical hierarchical routing protocols for wsns : A review. In International Conference on Computing, Communication and Automation (ICCCA), pages 465–470, 2016.

[27] A. Saeed, W. H. Butt, F. Kazmi, and M. A. Khan. Evaluation of clustering algorithms for wireless sensor and actor networks. In 15th International Bhurban Conference on Applied Sciences and Technology (IBCAST), pages 883–889, 2018.

[28] Sariga Arjunan and Sujatha Pothula. A survey on unequal clustering protocols in wireless sensor networks. Journal of King Saud University - Computer and Information Sciences, 31, 2017.

[29] Ali Shokouhi Rostami, Marzieh Badkoobe, Farahnaz Mohanna, Hengameh keshavarz, Ali Asghar Rahmani Hosseinabadi, and Arun Kumar Sangaiah. Survey on clustering in heterogeneous and homogeneous wireless sensor networks. The Journal of Supercomputing, 74(1):277–323, 2018.

[30] Muhammad Riaz. Clustering algorithms of wireless sensor networks : A survey. International Journal of Wireless and Microwave Technologies, 8:40–53, 2018.

[31] Rajesh D. and Giji Kiruba. A survey on energy efficient clustering algorithms in mobile wireless sensor networks. International Journal of Engineering and Robot Technology, 5:49–55, 2018.

[32] Fakhrosadat Fanian and Marjan Kuchaki Rafsanjani. Cluster-based routing protocols in wireless sensor networks: A survey based on methodology. Journal of Network and Computer Applications, 142:111 – 142, 2019.

[33] Olayinka O. Ogundile and Attahiru S. Alfa. A survey on an energy-efficient and energy-balanced routing protocol for wireless sensor networks. Sensors, 17, 2017.

[34] Y.Suresh, S.Roshni, J.Senthilkumar, and V.Mohanraj. Survey on algorithms for energy efficient cluster head selection in wireless sensor networks. In Journal of Information Systems and eBusiness Network, editors, Proceedings of the International Conference on Intelligent Computing Systems ICICS, pages 446–455, 2017.

[35] Er. Atul Goyal Priyanka Goel. Clustering in wireless sensor networks: Survey. International Journal of Advance Research and Innovative Ideas in Education, 3:361–365, 2017.

[36] A. Aalavandhar and A. Arjunan. A comparative study of different cluster mechanism for wireless sensor networks. In 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), pages 781–785, 2019.

[37] P. K. Mishra and S. K. Verma. A survey on clustering in wireless sensor network. In 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pages 1–5, 2020.

[38] Amin Shahraki, Amir Taherkordi, Oystein Haugen, and Frank Eliassen. Clustering objectives in wireless sensor networks: A survey and research direction analysis. Computer Networks, 180:107376, 2020.

[39] Djamel Djenouri, Nassima Merabtine, fatma zohra Mekahlia, and Messaoud Doudou. Fast distributed multi-hop relative time synchronization protocol and estimators for wireless sensor networks. Ad Hoc Networks, 11:2329–2344, 2013.

[40] Riham S.Y. Elhabyan and Mustapha C.E. Yagoub. Two-tier particle swarm optimization protocol for clustering and routing in wireless sensor network. Journal of Network and Computer Applications, 52:116 – 128, 2015.

[41] Sang H. Kang and Thinh Nguyen. Distance based thresholds for cluster head selection in wireless sensor networks. IEEE Communications Letters, 16(9):1396–1399, 2012.

[42] J. Lee and C. Teng. An enhanced hierarchical clustering approach for mobile sensor networks using fuzzy inference systems. IEEE Internet of Things Journal, 4(4):1095–1103, 2017.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3092509, IEEE Access

[43] J. Lee and W. Cheng. Fuzzy-logic-based clustering approach for wireless sensor networks using energy predication. IEEE Sensors Journal, 12(9):2891–2897, 2012.
[44] Hoda Taheri, Peyman Neamatollahi, Ossama Mohamed Younis, Shahrzad Naghibzadeh, and Mohammad Hossein Yaghmaee. An energy-aware distributed clustering protocol in wireless sensor networks using fuzzy logic. Ad Hoc Networks, 10(7):1469 – 1481, 2012.
[45] Ronald R. Yager. Fuzzy logic methods in recommender systems. Fuzzy Sets and Systems, 136(2):133 – 149, 2003.
[46] V. Pal, G. Singh, and R. P. Yadav. Network adaptive round-time clustering algorithm for wireless sensor networks. In International Conference on Advances in Computing, Communications and Informatics (ICACCI), pages 1299–1302, 2013.
[47] Y. Liao, H. Qi, and W. Li. Load-balanced clustering algorithm with distributed self-organization for wireless sensor networks. IEEE Sensors Journal, 13(5):1498–1506, 2013.
[48] Asgarali Bouyer, Abdolreza Hatamlou, and Mohammad Masdari. A new approach for decreasing energy in wireless sensor networks with hybrid leach protocol and fuzzy c-means algorithm. Int. J. Commun. Netw. Distrib. Syst., 14(4):400–412, 2015.
[49] S. A. Alharthi and P. A. Johnson. Threshold sensitive heterogeneous leach protocol for wireless sensor networks. In 2016 24th Telecommunications Forum (TELFOR), pages 1–4, 2016.
[50] S. Radhika and P. Rangarajan. On improving the lifespan of wireless sensor networks with fuzzy based clustering and machine learning based data reduction. Applied Soft Computing, 83:105610, 2019.
[51] Seyyit Alper Sert, Hakan Bagci, and Adnan Yazici. Mofca: Multi-objective fuzzy clustering algorithm for wireless sensor networks. Applied Soft Computing, 30:151 – 165, 2015.
[52] S. A. Sert, A. Alchihabi, and A. Yazici. A two-tier distributed fuzzy logic based protocol for efficient data aggregation in multihop wireless sensor networks. IEEE Transactions on Fuzzy Systems, 26(6):3615–3629, 2018.
[53] El Alami Hassan and Abdellah Najid. Energy-efficient fuzzy logic cluster head selection in wireless sensor networks. pages 1–7, 03 2016.
[54] V. Chauhan and S. Soni. Energy aware unequal clustering algorithm with multi-hop routing via low degree relay nodes for wireless sensor networks. J Ambient Intell Human Comput, 2021.
[55] Nabajyoti Mazumdar, Amitava Nag, and Sukumar Nandi. Hdds: Hierarchical data dissemination strategy for energy optimization in dynamic wireless sensor network under harsh environments. Ad Hoc Networks, 111:102348, 2021.
[56] M. Adnan, L. Yang, T. Ahmad, and Y. Tao. An unequally clustered multihop routing protocol based on fuzzy logic for wireless sensor networks. IEEE Access, pages 1–1, 2021.
[57] H. El Alami and A. Najid. Ech: An enhanced clustering hierarchy approach to maximize lifetime of wireless sensor networks. IEEE Access, 7:107142–107153, 2019.
[58] J. m. Kim, H. k. Joo, S. s. Hong, W. h. Ahn, and H. b. Ryou. An efficient clustering scheme through estimate in centralized hierarchical routing protocol. In International Conference on Hybrid Information Technology, volume 2, pages 145–152, 2006.
[59] A. Azim and M. M. Islam. A dynamic round-time based fixed low energy adaptive clustering hierarchy for wireless sensor networks. In 9th International Conference on Communications (MICC), pages 922–926, 2009.
[60] R. Goudarzi, B. Jedari, and M. Sabaei. An efficient clustering algorithm using evolutionary hmm in wireless sensor networks. In 8th International Conference on Embedded and Ubiquitous Computing (EUC) IEEE/IFIP, pages 403–409, 2010.
[61] B. Qi and F. Shen. Performance comparison of partical swarm optimization variant models. In 11th International Conference on Information Technology: New Generations (ITNG), pages 575–580, 2014.
[62] Chang jiang JIANG, Wei ren SHI, min XIANG, and Xian lun TANG. Energy-balanced unequal clustering protocol for wireless sensor networks. The Journal of China Universities of Posts and Telecommunications, 17(4):94 – 99, 2010.
[63] A. Patra and S. Chouhan. An approach to improved energy efficient hybrid clustering in wireless sensor networks. In International Conference on Signal Processing and Communications (SPCOM), pages 1–6, 2014.
[64] D. C. Hoang, P. Yadav, R. Kumar, and S. K. Panda. Real-time implementation of a harmony search algorithm-based clustering protocol for energy-efficient wireless sensor networks. IEEE Transactions on Industrial Informatics, 10(1):774–783, 2014.
[65] M. Hatamian, S. S. Ahmadpoor, S. Berenjian, B. Razeghi, and H. Barati. A centralized evolutionary clustering protocol for wireless sensor networks. In 6th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pages 1–6, 2015.
[66] Shayesteh Tabatabaei, Amir Rajaei, and Amir Mohsen Rigi. A novel energy-aware clustering method via lion pride optimizer algorithm (lpo) and fuzzy logic in wireless sensor networks (wsns). Wireless Personal Communications, 108(3):1803–1825, 2019.
[67] B. Meenakshi and P. Anandhakumar. Lifetime extension of wireless sensor network by selecting two cluster heads and hierarchical routing. In Proceedings of the International Conference on Advances in Computing, Communications and Informatics, ICACCI '12, pages 1254–1260, New York, NY, USA, 2012.
[68] Anar Abdel-Hady, Hossam M. A. Fahmy, Sherine M. Abd El-Kader, Hussein S. Eissa, and Ashraf Salem. Multilevel minimised delay clustering protocol for wireless sensor networks. Int. J. Commun. Netw. Distrib. Syst., 13(2):187–220, 2014.
[69] El Alami Hassan and Abdellah Najid. Ms-routing-gi: Routing technique to minimize energy consumption and packet loss in wsns with mobile sink. IET Networks, 7, 04 2018.

NASSIMA MERABTINE obtained the PhD in Computer Science from the Computer Science Higher School (ESI), Algiers, Algeria, in 2020. From 2014 to 2017, she worked as a full-time research support engineer at the Center for Development of Advanced Technologies (CDTA). She also served as a temporary assistant lecturer at the computer science department of Blida University during 2012 and 2013.

DJAMEL DJENOURI obtained the PhD in Computer Science from the University of Science and Technology (USTHB), Algiers, Algeria, in 2007. In 2008, he was granted a post-doctoral fellowship from the European Research Consortium on Informatics and Mathematics (ERCIM), and has been working at the Norwegian University of Science and Technology (NTNU), Norway, from 2008 to 2009. He was a senior research scientist (Director of Research) and deputy director at the CERIST, Algiers, and adjunct full professor at the EMP polytechnic university, Algiers. He is recently joined the University of the West of England (UWE), UK. He is working on topics related Internet of things, wireless and mobile networks, machine learning and application for smart cities and green applications. He has been conducting several research projects with international collaborations as the principal investor for many of them. He participated in many international conferences worldwide and gave many keynotes and plenary-session talks. He has been granted mobility internships for short visits to many renowned universities including NTNU, SICS (Stockholm), university of Cape town, UPC Barcelona, JMU Liverpool, University of Padova, the university of Oxford, etc. He has been a visiting researcher to NTNU in 2017 and 2019. He published more than 100 papers in international peer-reviewed journals and conference proceedings, two books, and he is holding two national patents. He voluntarily contributed to the organization of many conferences and workshops, and he served as TPC member of many international conferences, guest editor of a special issue, and reviewer for many Journals. He is a senior member of Association of Commuting Machinery (ACM), member of the Arab/German Young Academy of Science and Humanities (AGYA).

DJAMEL-EDDINE ZEGOUR is a professor of computer science at ESI Graduate School of Computer Science in Algiers. His main research interests include data structures and algorithms, compilers, functional programming, object-oriented programming and distributed computing.

• • •