A Framework of Modeling Large-Scale Wireless Sensor Networks for Big Data Collection

Asside Christian Djedouboum [1,2,3] , **Ado Adamou Abba Ari** [1,2,*] ,
Abdelhak Mourad Gueroui [1] , **Alidou Mohamadou** [2] , **Ousmane Thiare** [4]
and Zibouda Aliouat [5]

1   LI-PaRAD Laboratory, Université Paris Saclay, Versailles Saint-Quentin-en-Yvelines University,
    45 Avenue des États-Unis, 78000 Versailles, France; asside.djedouboum@univ-maroua.cm (A.C.D.);
    mourad.gueroui@uvsq.fr (A.M.G.)
2   LaRI Laboratory, University of Maroua, P.O. Box 814 Maroua, Cameroon;
    alidou.mohamadou@univ-maroua.cm
3   Faculty of Exact and Applied Sciences, University of Moundou, P.O. Box 206 Moundou, Chad
4   LANI Laboratory, Gaston Berger University of Saint-Louis, P.O. Box 234 Saint-Louis, Senegal;
    othiare@ugb.edu.sn
5   LRSD Laboratory, University Ferhat Abbes Setif 1, El Bez, Setif 19000, Algeria; zaliouat@univ-setif.dz
*   Correspondence: adoadamou.abbaari@gmail.com; Tel.: +33-1-39-25-40-67

check for updates

Abstract: Large Scale Wireless Sensor Networks (LS-WSNs) are Wireless Sensor Networks (WSNs) composed of an impressive number of sensors, with inherent detection and processing capabilities, to be deployed over large areas of interest. The deployment of a very large number of diverse or similar sensors is certainly a common practice that aims to overcome frequent sensor failures and avoid any human intervention to replace them or recharge their batteries, to ensure the reliability of the network. However, in practice, the complexity of LS-WSNs pose significant challenges to ensuring quality communications in terms of symmetry of radio links and maximizing network life. In recent years, most of the proposed LS-WSN deployment techniques aim either to maximize network connectivity, increase coverage of the area of interest or, of course, extend network life. Few studies have considered the choice of a good LS-WSN deployment strategy as a solution for both connectivity and energy consumption efficiency. In this paper, we designed a LS-WSN as a tool for collecting big data generated by smart cities. The intrinsic characteristics of big data require the use of heterogeneous sensors. Furthermore, in order to build a heterogeneous LS-WSN, our scientific contributions include a model of quantifying the kinds of sensors in the network and the multi-level architecture for LS-WSN deployment, which relies on clustering for the big data collection. The results simulations show that our proposed LS-WSN architecture is better than some well known WSN protocols in the literature including Low Energy Adaptive Clustering Hierarchy (LEACH), E-LEACH, SEP, DEEC, EECDA, DSCHE and BEENISH.

Keywords: big data; large-scale wireless sensor network; clustering; data collection; framework; modeling

1. Introduction

The last decade has undeniably been the decade of the rapid growth of wireless communication technologies [1,2]. However, the many application perspectives of wireless communication-based applications, including increasingly common Wireless Sensor Networks (WSNs), continue to pose major technical and scientific challenges [3,4]. Services and applications based on WSNs require a

communication infrastructure whose performance must be continuously studied and improved in order to better adapt to the constraints and quality of service requirements of new applications and to the increased use of these networks.

In recent years, the massification of digital data has been almost exponential. Moreover, the Statista (Statista Research Department. Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025 (in billions). https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/) portal (one of the world's largest portals for statistics and market data) goes even further by estimating that there would currently be 26.66 billion connected objects compared to 15.41 billion in 2015, and that by 2025, it estimates that the number of connected objects in circulation worldwide will amount to 75.44 billion. Figure 1 perfectly illustrates the correlation between the number of connected objects and the amount of massive data generated during these years. The massive data or the big data is considered by many researchers as one of the major challenges of modern computing in the current decade. This mass of data poses serious problems, since it is difficult or even impossible to capture and process this data using traditional data processing tools. Certainly, the cloud enabled a number of computing and storage features which benefits to big data based applications [5].
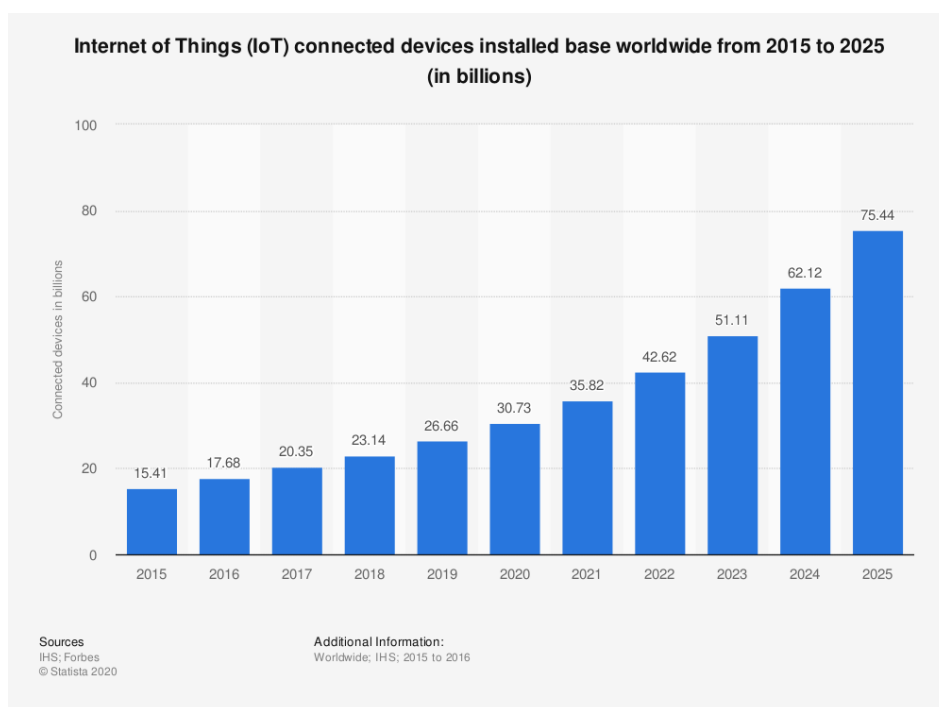

Figure 1. This statistic represents the number of connected devices (Internet of Things—IoT) in the world between 2015 and 2025. In 2019, the source predicted that 26 billion connected objects would be in circulation worldwide [6].

This massive or big data is considered by many researchers as one of the great challenges of modern computing in the current decade. This mass of data poses serious problems, as it is difficult if not impossible to capture and process this data using traditional data processing tools. Furthermore, the number of application perspectives of WSNs, including precision agriculture, forest monitoring for fire detection, patient monitoring, natural disaster management, etc., makes it possible to consider the use of these networks for collecting big data generated by smart cities [7–9]. However, in such contexts, these WSNs are consisting of an almost large number of sensors to be deployed over large areas [1]. Deploying such a large number of sensors is a common practice to overcome frequent sensor failures and avoid human intervention to replace them or recharge their batteries. This is a solution to ensure a reliable network that can last over time considering the spatial redundancy of

the sensors. In reality however, high sensor density can be a major waste of energy and resources if coupled with a poor deployment strategy and lack of good communication organization and routing protocol. In addition, high density can lead to a large number of collisions and interference, leading to over-consumption of energy on necessary retransmissions due to packet losses, and consequently, a loss in overall performance (significant delays and packet losses) [10]. Therefore, proposing a big data collection scheme that extend the battery lifetime is an important issue [11].

Besides, in the Large Scale Wireless Sensor Network (LS-WSN) applications, the sensor deployment strategy has a strong impact on the quality of communications. Indeed, since the communication range of the sensors is limited, a random deployment of the sensors can lead to connectivity and coverage gaps. In addition, a poor deployment strategy can lead to unbalanced energy depletion, resulting in empty areas over time while others remain quite dense. Therewith, in the literature, the WSN deployment techniques proposed so far have in most cases been designed to either maximize network connectivity, increase coverage of the area of interest or, of course, extend the network's lifespan [12–14].

However, few studies in the literature have considered the choice of a good sensor deployment architecture as a solution for both connectivity and energy optimization in Large-Scale Wireless Sensor Networks (LS-WSNs) [1,15,16]. The different deployment architectures, whether deterministic or random, generally consider a single objective to be achieved, that of coverage or connectivity. In addition, most energy conservation deployment approaches consider the uniform redundancy of sensors over the area of interest and the healing of connectivity gaps as the only objective. No strategy, to the best of our knowledge, considers the imbalance in energy consumption due to communications and its link with routing. Hence, we are positioning ourselves as part of the deployment of a LS-WSN in which we propose an optimal architecture for deployment and data routing that must: (i) ensure optimal connectivity and coverage of the area of interest, (ii) minimize the energy consumption of the sensors, (iii) extend the life of the sensors and the network in general, and (iv) adjust the network topology following a connectivity failure due to various sensor failures.
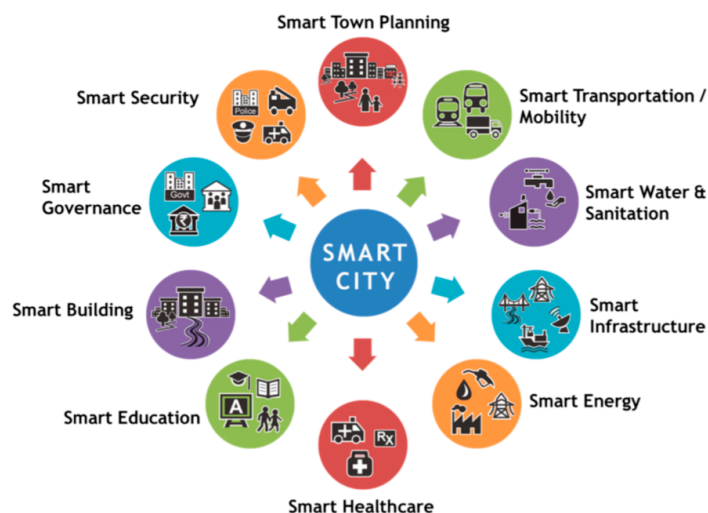

Figure 2. The big data application perspectives in a smart city.

As illustrated in Figure 2, the big data presents many application perspectives for smart cities. Unfortunately, the issue of collecting and processing this massive data remains a challenge for computer science research. We propose to use LS-WSNs to address the challenges of collecting these data. The intrinsic characteristics impose the use of heterogeneous sensors. For this purpose, our main contributions range from the proposal of a mathematical model, which from a predefined level of heterogeneity, determines the number of the different sensors and the amount of energy related to them for the construction of the network. We have opted to build the network according to a multi-level hierarchical architecture where the sensors will be organized in clusters. Thus, we have proposed

a clustering algorithm that best fits the scaling of the sensors. Finally, the different contributions are simulated under OMNET++ coupled with the INET Framework [17]. In summary, our main contributions are structured around the following points:

- Proposal of a computation model that determines a set of sensors N and the level of heterogeneity α as well as the respective number of the different types of sensors to be used;
- Proposal for a multi-level architecture of LS-WSN that optimizes connectivity and the sensor's energy consumption;
- Implementation of an algorithm for building clusters of our architecture;
- Proposal for a pre-established routing mechanism in which routing paths are less costly in terms of power consumption;
- Simulation of our proposed LS-WSN model.

The rest of the article is organized as follows: Section 2 presents some related works big data related to WSNs. This makes it possible to deduce an architecture that best meets the challenges of big data collection. The main contributions of the paper are presented in Sections 3 and 4; The results of the performance evaluation are discussed in Section 5. The conclusion and future research are presented in Section 6.

2. Big Data, Dimensions and Analysis Tools Related to WSNs

Big data characterizes the set of large volumes of data for which it is difficult or even impossible to collect and process using traditional data processing tools. The literature defines big data according to a formalism called V; three to five Vs allow the characterization of this mass of data [18]. For Doug Laney [1], big data is characterized by the volume, speed, and variety of data, giving rise to the 3 Vs principle. Volume describes the size of the data, velocity refers to the speed at which the data is produced while variety describes the range of data types and sources. Recently, an additional "V" has been added to this definition. For example, in [1,19], the authors have added two "Vs" to the first three. Thus, the fourth "Vs" refers to the value or variability while the fifth "Vs" refers to the veracity of data. Other more recent works have further integrated other Vs (6 Vs, 7 Vs, and 9 Vs) [20] to further define the contours of these big data. Figure 3 shows the main large dimensions of the data.
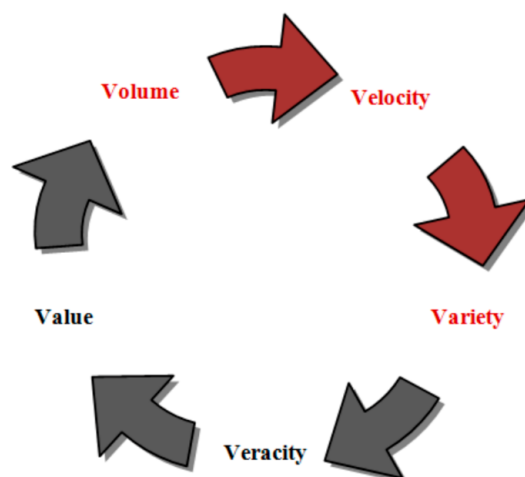

Figure 3. 5V characterizing the big data.

The problem of big data has led to a lot of research in the recent years [21,22]. In the context of WSNs, most works process and analyses big data from these networks. Techniques and algorithms based on Hadoop and MapReduce technology proposed by Web giants [23] are implemented in most cases. Therefore, in [24], the authors propose a set of tools for analyzing data collected by WSNs. Especially, they exploited the Hadoop data warehouse framework as well as the Hadoop virtual cluster

to design their data warehouse protocol, namely Hive [25]. The proposal also has a module called Hive Query Language (HiveQL) exploiting the Structured Query Language (SQL). The HiveQL requests are converted into MapReduce jobs. On the other hand, the work of [26] has integrated large data analysis tools into pollution monitoring sensors to collect, store and process data captured by this vast network. To do this, authors proposed a two modules based model: a data acquisition module (DAM) for data collection, and a data pre-processing, processing and analysis module (DPM) for real-time detections.

On the other hand, in [24], for processing large data while saving the energy consumption in a distributed wireless sensor network, the authors designed a data aggregation technique based on the Hadoop framework with simple/multi-cluster architectures. Hence, to the best of our knowledge, there is little work that deals with the collection of big data using LS-WSN [1]. This issue is an interesting challenge on which we are positioning ourselves. Given the intrinsic characteristics of big data, we have chosen to use a set of heterogeneous sensors [27]. Deploying this number of heterogeneous sensors requires a deployment and structuring strategy. We propose a multi-level architecture based on clustering, which includes a model for quantifying the different types of sensors that make up the network.

Furthermore, LS-WSNs are constraining networks (lack of infrastructure, resource constraints, heterogeneity, and network dynamics). Therefore, it is important to think about a self-organized, adaptive, and energy-efficient virtual topology. To design such a topology, several solutions have been proposed in the literature such as clustering and backbone [28]. Several techniques have been proposed to significantly increase the lifetime of cluster-based networks by partitioning the network into groups such that the intergroup distance is less than the extra-group distance [4,29]. Each network group is managed by a Cluster Head (CH). The choice of a CH is either the result of an elective process where the nature of a sensor node predisposes it to this role or the role is fixed in a centralized manner [30,31]. In the case of clustering where the choice of the CH is based on an election is divided into two parts: one where the CH designation metrics do not take into account the energy of the candidate sensor nodes and the other where the CH designation metrics take into account the energy of the candidate nodes. Indeed, the first case where the choice of the CH does not take energy into account while in the second category, several algorithms place particular emphasis on the energy of sensor nodes that are candidates for the CH features. One of the most popular is the Low Energy Adaptive Clustering Hierarchy (LEACH) algorithm proposed in [32].

3. Model for Quantifying the Sensors of a Heterogeneous LS-WSN

Our LS-WSN model consists of heterogeneous sensors. For N sensors to be deployed and at a level α of heterogeneity, i.e., the number of different kinds of sensors, we propose to determine the number of different types of sensors involved in the formation of the network.

3.1. Network Assumptions

We adopted the following assumptions:

- Initially, all wireless sensors have the same characteristics instead of the energy supply that is different from a wireless sensor to another. Moreover, each wireless sensor is identified by a unique identifier ID and it is assumed that all sensors are stationary after the network deployments.
- The WSN is heterogeneous.
- The sensors do not know their location, i.e., they are not equipped with a GPS or an antenna.
- The sensors are left unattended after deployment, which means that it is impossible to recharge the sensor's battery.
- There is a unique stationary base station (BS) that has a stable power supply.
- Each CH performs data aggregation.
- The distances among the sensors are calculated on the basis of the received signal strength. Indeed, when travel toward the receiver, the transmitted signal is attenuated. According to Farooq-i-Azam

and Ayyaz [33], this distance is calculated according to transmitted power signal by the sender sensor, the strength of received power of the signal, and the path loss. More generally, distance calculation based on Received Signal Strenght Indicator (RSSI) saves power and no need to add additional circuits in the sensor device.

- The sensors have the ability to control the transmission energy as a function of the distance from the receiving nodes. The node failure is due to energy depletion. In fact, if the transmission distance is too large, the energy used for the transmission of one bit information is enough. Therefore, instead of transmitting data to a far sensor, a given sensor will prefer to transmit to a near sensor and the last will transmit to near neighbor in the same way until reaching the destination sensor that is far from the sender sensor.

- The energy consumption of the data transmission as well as data reception are similar. This is favored by the wireless radio link.

- Sensors randomly equipped in the monitoring area and nodes are indirectly managed by the BS. In fact, according the three-tier architecture presented hereinafter in Figure 5, sensor nodes are led by the cluster heads and the last are managed by the BS.

- Dead sensor IDs are not reused for other sensors.

3.2. Energy Consumption Model

As described in [30], our energy model uses sensors embedded with the realistic characteristics of the Chipcon CC2420, the radio transceiver whose datasheet is given in [34]. The CC2420 Chipcon is a radio transceiver inline with the IEEE 802.15.4 GHz to 2.4 GHz standard and complies with the ZigBeeTM standard, designed for less energy based WSN applications. CC2420 characteristics include multiple transmissions, hardware support for packet processing, data buffering, clear channel evaluation, link quality indication, and packet synchronization.

In accordance with the implementation of the IEEE 802.11 Received Signal Strenght Indicator (RSSI) that is a measure of the power available in a received radio signal, the power of the received signal allows to quantify the power consumption in a WSN environment [35]. To exploit this in our energy quantification, we assume that a sensor battery has linear discharge and charge features. Thus, the energy E_i consumed by the sensor i is equivalent to the sum of used energy resulting from that of its components [36]. The energy consumption of the components contains the energy used to execute events and the energy used in the transition between states. Therefore, the total energy consumed by the sensor i, E_i is given by Equation (1).

$$E_i = \sum (E_S + E_T) = \sum_j (p_j \times t_j) + \sum_T (p_T \times t_T)) \tag{1}$$

where E_S is the energy expended by a sensor inside the states: the index j refers to one of the four states of the CC2420 Chipcon (inactive, standby, receive or transmit). p_j is the average of the power consumed in each state j; and t_j is the operating time in the corresponding state. Then, E_T is the energy spent in transition between states: p_T is the average of the power consumed in transition state T and t_T is the operating time during the state transition.

3.3. Network Coverage Model

Coverage is a very important performance measure in WSN [15]. There are several types of coverage in WSN: point coverage, surface coverage, area coverage, and barrier coverage. However, we consider point, surface, and region coverage in our study because these three types of coverage are more than sufficient to study the coverage properties of most WSN applications.

- Coverage of points in the WSN. Let S be a given area of interest to be monitored. It is said that a sensor N_i covers a point $s \in S$, if and only if:

$$d(N_i, s) \leq R \tag{2}$$

where R is the communication range characterizing each node and $d(u, v)$ defines the Euclidean distance between the nodes u and v.

A point $s \in S$ is said to be k-covered by a set of k sensors $N_1, N_2, \cdots, N_k$ if and only if each of these k sensors covers both the point s, i.e., if and only if:

$$\forall N_i \in \{N_1, N_2, \cdots, N_k\}, d(N_i, s) \leq R \tag{3}$$

- Surface coverage in the WSN. The coverage of the surface of an area of interest by a sensor N_i is defined as the total area within the detection range of N_i. Analytically, a surface coverage by a sensor node N_i noted $C(N_i)$ is defined by the formula given in Equation (4).

$$C(N_i) = \{s \in S | d(N_i, s) \leq R\} \tag{4}$$

- Regional coverage in the WSN. Either A a region (zone) or s any point of S. The coverage of region A by a set of sensors $M_N = \{N_1, N_2, \cdots, N_N\}$ is defined analytically by:

$$\forall q \in A, \exists S_k \in M_N | d(N_i, s) \leq R \tag{5}$$

3.4. Multilevel Heterogeneous Network Model for LS-WSN

The different kinds of sensors and the associated energy resources can be quantified using a mathematical model. Less models in the literature take into account the general heterogeneity at several levels. The number of sensors in a network and their energy resources are completely independent, which is not, for instance, the case in the works of Quig et al. [37], since the authors randomly assign each sensor an energy source for a given interval. Inspired by the model described in [38] for WSNs, we designed a generic model for LS-WSNs. A flowchart of the proposed model is given in Figure 4.
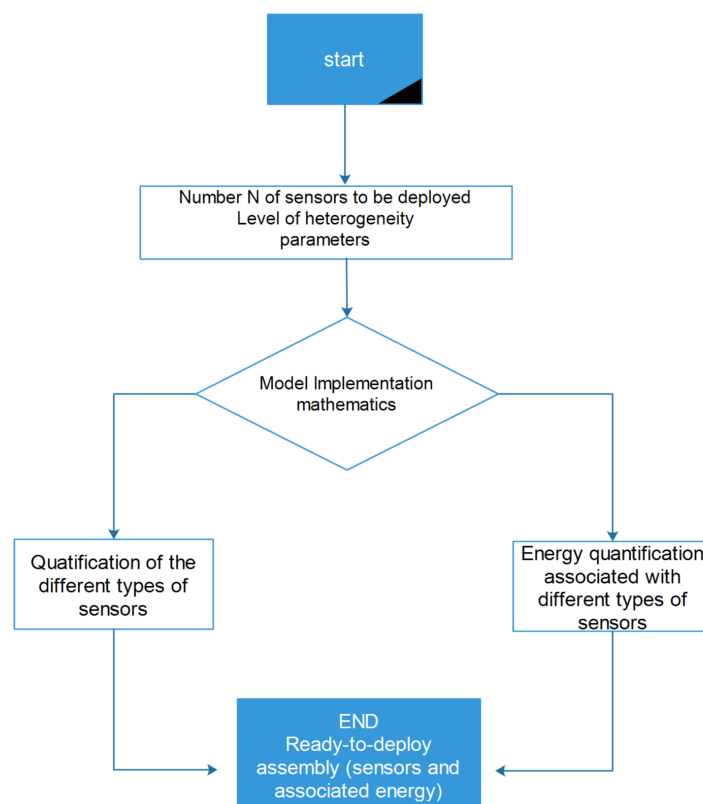

Figure 4. Flowchart of the model.

Then, let N be the total number of nodes of a network with n determining the level of heterogeneity. Note that the level of heterogeneity is the number of sensors from different components of the network. The total number N of sensors can, therefore, be divided according to the n node types, i.e., type-1, type-2, type-3,$\cdots$ type-n nodes with their respective energies as $E_1, E_2, E_3, \cdots, E_n$. The secondary parameters used in the model are determined by the value of n. In other words, for describing level-n heterogeneity, the network model should have n secondary parameters. Therefore, the energy levels must satisfy the condition given in Equation (6).

$$E_1 < E_2 < E_3 < E_4 < E_5 < \cdots < E_n \tag{6}$$

$N_1, N_2, N_3, \cdots, N_n$ that represents respectively, the numbers of type-1, type-2, type-3, $\cdots$, type-n nodes in the network, must satisfy the inequalities given in Equation (7).

$$N_1 < N_2 < N_3 < \cdots < N_n \tag{7}$$

The energy of the different types of sensors in the network is linked by the relationship given in Equation (8).

$$E_j = E_1 \times (1 + (j - 1) \times \delta) \tag{8}$$

where, E_1 is the energy of a type-1 of the sensor and $E_j, j = \{1, 2, 3, \cdots, n\}$, the energy of a type-$j$ of the sensor. The energy of a type-j sensor is δ times more than that of a type-$(j - 1)$ sensor, δ is a constant.

Then, the overall energy consumed in the network is given in Equation (9).

$$\begin{aligned} E_{total} = N \times \Big((\alpha - \beta_1) \times E_1 + (\alpha - \beta_1) \times (\alpha - \beta_2) \times E_2 \\ + (\alpha - \beta_1) \times (\alpha - \beta_2) \times (\alpha - \beta_3) \times E_3 + \\ \cdots + (\alpha - \beta_1) \times (\alpha - \beta_2) \times (\alpha - \beta_3) \times \cdots \times (\alpha - \beta_n) \times E_n \Big) \end{aligned} \tag{9}$$

α is the primary parameter in the model given in Equation (9). α determines the heterogeneity level of the overall network and the last is related to $\beta_i, i = 1, 2, \cdots, n$ by Equation (10).

$$\Big((\alpha - \beta_1) \times (1 + (\alpha - \beta_2) \times (1 + (\alpha - \beta_3) \times \cdots \times (1 + (\alpha - \beta_{n_1}))) \Big) = 1 \tag{10}$$

Then let give β_i given in Equation (12) be the secondary parameters, such a way that the relation given in Equation (11) is always true.

$$(\alpha - \beta_i) < 1 \tag{11}$$

$$\beta_i = \beta_{i-1} - 2 \times \gamma \tag{12}$$

where γ is a constant that is upperbounded for level-n of heterogeneity given in Equation (13).

$$\frac{\beta_i}{2(n - 1)} > \gamma \tag{13}$$

Then, if α is assigned the value $\beta_i (i > 1)$, i.e., $\alpha = \beta_i$, we find that $(i - 1)$ non-zero terms according to Equation (9). This means that there are only $(i - 1)$ types of sensor nodes in the network and that the model described by $(i - 1)$ the heterogeneity level.

For $i = 1$, the value of the model given in Equation (9) is nil. This does not mean any level of heterogeneity, it is the degenerative case.

For $\alpha = \beta_2$, we deduce that there is only node of type-1 in the network, which is actually a homogeneous network. However, the model describes a level-n heterogeneous network. According to

the model given in Equation (9), we deduce the energy of the heterogeneous network (level-1) by the formula given in Equation (14).

$$E_{level-1} = N \times (\alpha - \beta_1) \times E_1 \tag{14}$$

Moreover, the number N_1, of nodes of type-1 is given in Equation (15).

$$N_1 = N \times (\alpha - \beta_1) \tag{15}$$

According to Equation (10), we have N_1 equal to N since $(\alpha - \beta_1) = 1$.

For $\alpha = \beta_3$, we have only two non-zero terms in Equation (9). In this case, the model describes a heterogeneous level-2 with a total energy $E_{level-2}$ given in Equation (16).

$$E_{level-2} = N \times \left((\alpha - \beta_1) \times E_1 + ((\alpha - \beta_1) \times (\alpha - \beta_2) \times E_2) \right) \tag{16}$$

The number N_1 (respectively N_2) of nodes of type-1 (respectively type-2) in the network is given in Equation (17) (respectively Equation (18)).

$$N_1 = N \times (\alpha - \beta_1) \tag{17}$$

$$N_2 = N \times \left((\alpha - \beta_1) \times (\alpha - \beta_2) \right) \tag{18}$$

According to Equation (10) we have $(\alpha - \beta_1) + (\alpha - \beta_1) \times (\alpha - \beta_2) = 1$.

For $\alpha = \beta_4$, we count three non-zero terms in the Equation (9) and in this case, the model describes a heterogeneous level-3 network. The total energy of the network is given in Equation (19).

$$E_{level-3} = N \times \Big((\alpha - \beta_1) \times E_1 + (\alpha - \beta_1) \times (\alpha - \beta_2) \times E_2 + $$
$$(\alpha - \beta_1) \times (\alpha - \beta_2) \times (\alpha - \beta_3) \times E_3 \Big) \tag{19}$$

The number N_3 of sensors of type-3 is given in Equation (20).

$$N_3 = N \times \left((\alpha - \beta_1) \times (\alpha - \beta_2) \times (\alpha - \beta_3) \right) \tag{20}$$

According to Equation (10) we have $\left((\alpha - \beta_1) \times (1 + (\alpha - \beta_2) \times (1 + (\alpha - \beta_3))) \right) = 1$.

For level-i of heterogeneity e.g., $\alpha = \beta_{i+1}$, we have a heterogeneous network at level-i, whose total energy $E_{level-i}$ is given in Equation (21).

$$E_{level-i} = N \times \Big((\alpha - \beta_1) \times E_1 + (\alpha - \beta_1) \times (\alpha - \beta_2) \times E_2 + $$
$$(\alpha - \beta_1) \times (\alpha - \beta_2) \times (\alpha - \beta_3) \times E_3 + $$
$$\cdots + (\alpha - \beta_1) \times (\alpha - \beta_2) \times \cdots \times (\alpha - \beta_i) \times E_i \Big) \tag{21}$$

More generally, the number N_i of sensors of type-i is given in Equation (22).

$$N_i = N \times \left((\alpha - \beta_1) \times (\alpha - \beta_2)) \times (\alpha - \beta_3) \times \cdots \times (\alpha - \beta_i) \right) \tag{22}$$

According to Equation (10), we have $\Big((\alpha - \beta_1) \times (1 + (\alpha - \beta_2) \times (1 + (\alpha - \beta_3))) \times \cdots \times (1 + (\alpha - \beta_{i-1})) \Big) = 1$.

Therefore, the network model above described is a generic multi-level heterogeneous network model that can describe any level of heterogeneity in the network.

4. Clustering Algorithm for LS-WSNs

In this section, we propose a clustering algorithm for LS-WSNs that aims at maximizing the connectivity and optimize energy consumption.

4.1. LS-WSN Architecture

The LS-WSN model we propose is built according to a three-level architecture (see Figure 5): the first level consists of a set of sensors (member nodes) whose role is to gather and send information to their corresponding CHs. The sensors do not have the same communication range.
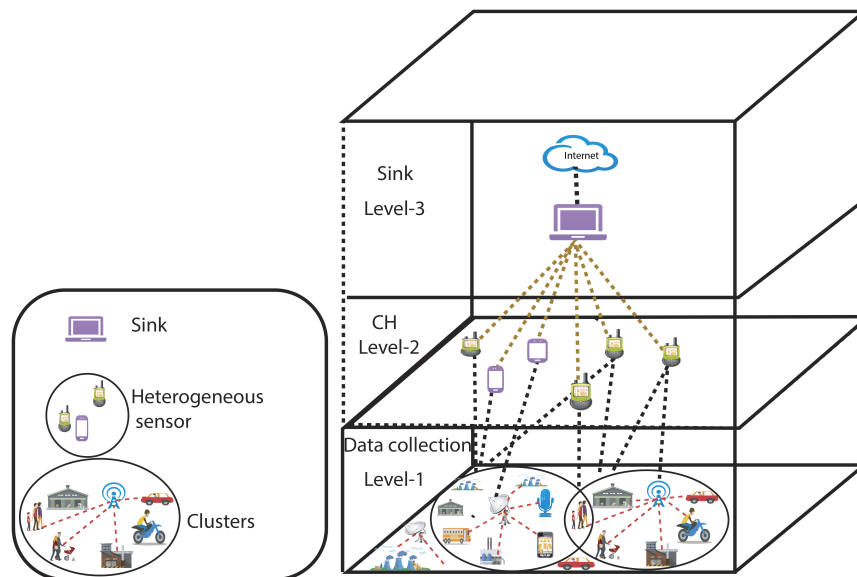

Figure 5. Three-tier architecture of LS-WSN.

All the CHs are in the second level of our architecture. CHs aggregate the data coming from their members and send the created packet to the BS. A cluster is led by a CH. The member nodes as well as the CHs have the same technical characteristics. The BS is the level-3, it processes the data received from the CHs according to predefined programs. Moreover, nodes that are members of a cluster communicate with their CHs (intra-cluster connectivity at 1-hop). In the same way, the CHs communicate with the BS. This communication procedure is the one defined in the LEACH clustering algorithm [39] whose purpose is to minimize the energy consumption of the nodes during communication. Furthermore, we assume that the first level sensor nodes work on 802.15.4 (zigbee) frequency channels. We also assume that the CHs use the protocol stack of the 802.11 standard for their communication with the BS.

4.2. Cluster Building Algorithm

We model our network by a graph $G = (V, E)$ where V represents all sensor nodes and $E = \{(u, v) \in V^2 \mid d(u, v) \leqslant R\}$ represents all wireless links among nodes. A wireless link exists between a pair of nodes (u, v) if they are within the communication range of each other. We offer a suitable and efficient clustering algorithm for LS-WSNs with multiple wells and channels. To build network clusters, we define four states for one node:

- Ordinary: initial state of a sensor disconnected from the communication structure.
- Leader: state of a well initiating the construction of its cluster. This is the root of the tree in formation or the CH.
- Member: intermediate node between the root and the leaves of a cluster tree.

- Gateway: intermediate node between clusters.

Our clustering algorithm builds cluster trees with *k*-hops (the distance between a node and its cluster leader is at most *k*-hops). The Cluster formation begins with a neighborhood discovery phase followed by the cluster construction phase initiated by the various cluster construction sensors. The different phases of our algorithm are summarized in the Figure 6 and described in the Algorithm 1. The proposed clustering is dynamic, so sensors can integrate or exit a cluster at any time. The procedure for a sensor to join a new cluster is described as follows:

- Procedure to join a new cluster: The BS periodically sends the list of CHs as well as their locations to the nodes disconnected from the structure (not belonging to a cluster). Each node calculates at each period its distances to the different CHs, if a distance will be *R* then it sends a "hello" message to the CH concerned. The CH sends him his ID and then the sensor joins the cluster by sending back a message "*clusterhead_accepted*".
- Sending information to the BS: Each member node has a standby period *T*. He wakes up every time, picks up information and sends it to his CH. The CH aggregates the received information and sends the built message to the BS.
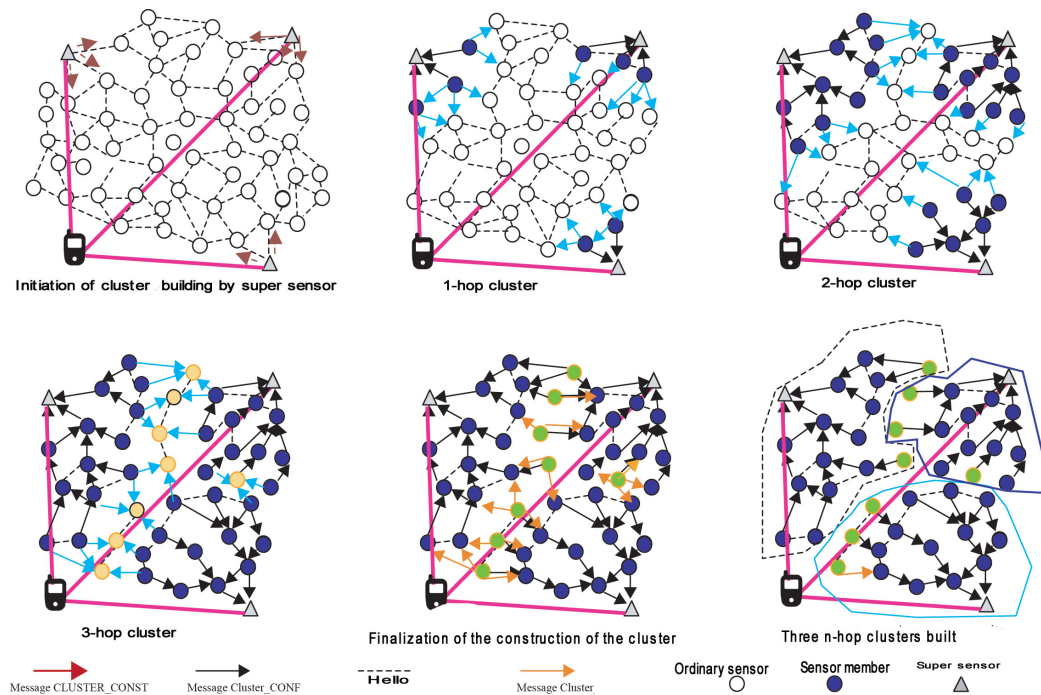

Figure 6. Overview of the cluster formation process according to our heuristics.

The implementation of our proposed algorithm is subject to a number of specific conditions that are:

1. Condition 1: A node receiving two "*CLUSTER_CONF*" messages from the two CHs, chooses the one with the lowest weight.
2. Condition 2: A node with a degree of zero (not having neighbors), sends its data directly to the BS and triggers the procedure to join a new cluster'.
3. Condition 3: A node leaving its cluster, sends its data to the BS and triggers the procedure to join a new cluster.

Algorithm 1 Large Scale Wireless Sensor Network (LS-WSN) clustering algorithm

Step: 1

After deployment, each sensor sends a "hello" message to the other sensors to allow it to discover its 1-hop neighborhood.

Step: 2

Creation and update of the neighborhood table of each sensor node after receiving the "hello" message from its neighborhood.

Step: 3

For cluster creation, each CH sends a message "$CLUSTER\_CONST$" to the sensor nodes, inviting them to join the cluster they want to build.

Step: 4

Upon receipt of the invitations, each sensor node responds with a procedure that we call $reception\_CLUSTER\_CONST$. This procedure updates the neighborhood table of the ordinary node and decides whether or not to integrate the cluster under construction.

Step: 5

Acceptance of the CH invitation. The sensor node becomes a member of the cluster after accepting the message from a CH. The latter then issues a confirmation message $CLUSTER\_CONF$ at 1-hop to notify the cluster of its membership and to invite its 1-hop neighbors to join it if they have not yet joined a cluster.

Step: 6

Upon receipt of the message $CLUSTER\_CONF$, depending on the nature of the sensors (CH or sensor node) a $CLUSTER\_CONF$ procedure is executed.

- If the receiver sensor of the $reception\_CLUSTER\_CONF$ is a CH, it updates its neighborhood table and then stops the retransmission of the $CLUSTER\_CONST$ message to the other sensors. It adapts its transceiver channel to the radio channel allocated to the cluster.
- In case the sensor is already a member of the cluster. The latter updates his table and chooses as a father the node with the best weight between this node and his current father. Then, he stops the retransmission of his message $CLUSTER\_CONF$. If it has not yet done so and is closer to the CH than the ordinary sensor, then it physically adapts its transceiver to the radio channel allocated to the cluster.
- If the sensor is not a member of a cluster, then the cluster updates its neighborhood table, then chooses as a father among its neighbors members to a jump, the one who has the best weight. On the other hand, if the sensor receives more than one $CLUSTER\_CONF$ message, then it becomes the gateway and issues a $CLUSTER\_END$ message, otherwise it becomes a member and issues the $CLUSTER\_CONF$ message to notify its cluster membership and invite others to join.

Step: 7

The management of the $CLUSTER\_END$ procedure. Upon receipt of this message is executed the procedure called "$reception\_CLUSTER\_END$" depending on the nature of the sensors:

- If the sensor is a CH, the latter updates its neighborhood table.
- If the sensor is a member of a cluster. The latter updates his neighborhood table. If necessary, it stops the retransmission of the message $CLUSTER\_CONF$ and $CLUSTER\_END$, if it has not already done so, it physically affects its transceiver and the radio channel allocated to the cluster.

---

5. Results and Discussion

In this section, we evaluate the performance of our heterogeneous LS-WSN model dedicated to the collection of big data generated in smart cities. This section assesses the main contributions proposed in the manuscript. First of all, it is the model for calculating the various sensors proposed in Section 3. Then, the clustering algorithm dedicated to LS-WSN proposed in Section 4. The implementation of our solutions was done using the OMNET++ simulator coupled with the INET Framework. More specifically, through our performance evaluation, we sought to see the impact of the level of sensor heterogeneity on the performance in terms of lifetime and energy consumption of the network but also

in terms of data transmitted to the well, latency, etc. Then, we evaluate the impact of the scope of the sensors in the formation of clusters. A comparative study of our LS-WSN model with some recent solutions proposed in the literature have been done.

5.1. Evaluation of the Different Sensors of the LS-WSNs

We have already shown in Section 3 that our model can describe any level of heterogeneity of LS-WSN. For validation purposes, we illustrated this by implementing a LS-WSN with up to six levels of heterogeneity, i.e., LS-WSN that can have type-1, type-2, type-3, type-4, type-5 and type-6 of the sensor. We define LS-WSN-1, LS-WSN-2, LS-WSN-3, LS-WSN-4, LS-WSN-5 and LS-WSN-6, respectively, as level-$1, 2, 3, 4, 5$ and 6 of heterogeneous networks.

We consider 10,000 sensors to be deployed. For each type of sensor, we use our proposed calculation model to determine the number of different types of sensors and the associated energy respectively. The input parameters adopted by our model are provided in the Table 1.

Table 1. Parameters.

Description	Notation	Value
Number of sensor nodes	N	10,000
Number of super sensors	Ns	110
Initial Energy	E_1	0.2 J
Constancy	α, δ	0.5, 0.025
Model parameterization	β_1	0.4

The parameter value α of the model is crucial. In fact, it determines the level of heterogeneity of the network. Thus, for an α respectively equal to $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ and β_6, which defines a network with a heterogeneity level of $1, 2, 3, 4, 5$ and 6 respectively. On the other hand, the energy of the different types of sensors defined by our model is decreasing and satisfies the following inequality given in Equation (23).

$$E_6 > E_5 > E_4 > E_3 > E_2 > E_1 \tag{23}$$

Thus, for:

- $\alpha = \beta_2$, the model describes a level-1 heterogeneous network, i.e., it is a homogeneous network using the same kind of sensors. For this network, if we consider 10,000 sensors to deploy, these sensors will all be of the same type, or $10,000$ sensors of the same nature.
- $\alpha = \beta_3$, the model describes a level-2 heterogeneous network, it is a heterogeneous network with two types of sensors. For this network, if we consider 10,000 sensors to deploy, we will have 6000 sensors of type-1 and 4000 sensors of type-2.
- $\alpha = \beta_4$, the model describes a level-3 heterogeneous network, it is a heterogeneous network with three types of sensors. For this network, if we consider 10,000 sensors to deploy, we will have 5200 sensors of type-1, 3000 sensors of type-2 and 1800 sensor of type-3.
- $\alpha = \beta_5$, the model describes a level-4 heterogeneous network, it is a heterogeneous network with four types of sensors. For this network, if we consider 10,000 sensors to deploy, the sensors of type-1, type-2, type-3, and type-4 are respectively 4900, 2600, 1500, and 1000 sensors.
- $\alpha = \beta_6$, the model describes a level-5 heterogeneous network, it is a heterogeneous network with four types of sensors. For this network, if we consider 10,000 sensors to deploy, the sensors of type-1, type-2, type-3, type-4, and type-5 are respectively 4700, 2400, 1400, 900, and 600 sensors.
- $\alpha = \beta_7$, the model describes a level-6 heterogeneous network, it is a heterogeneous network with four types of sensors. For this network, if we consider 10,000 sensors to deploy, the sensors of type-1, type-2, type-3, type-4, type-5, and type-6 are respectively 4608, 2354, 1320, 806, 533 and 378 sensors.

We will agree that the categorization of the number of sensors is not random. It is determined through the equations proposed by our model, in particular the formula given in Equation (16). On the other hand, the energy associated with the different types of sensors is obtained in Equation (9).

Therefore, for an:

- LS-WSN-1, i.e., a homogeneous network consisting of a set of sensors of the same type, the initial energy of each of these sensors is 0.2 J.
- LS-WSN-2, i.e., a heterogeneous network with two types of sensors (type-1, type-2). The energy of these different types is 0.2 J and 0.4 J respectively.
- LS-WSN-3, i.e., a heterogeneous network with three types of sensors: type-1, type-2, type-3, their energy is respectively 0.2 J, 0.4 J and 0.6 J.
- LS-WSN-4, i.e., a heterogeneous network with four types of sensors. The energy of sensors of type-1, type-2, type-3 and type-4 is 0.2 J, 0.4 J, 0.6 J and 0.8 J respectively.
- LS-WSN-5, i.e., a heterogeneous network with five types of sensors. The energy of the sensors of $type_1$, $type_2$, $type_3$, type-4, and type-5 are respectively 0.2 J, 0.4 J, 0.6 J, 0.8 J and 10 J.

Illustrative example:

For the purpose of demonstrating the enumeration of the different sensor types through the proposed model, we chose a level of heterogeneity equal to 6, always considering the 10,000 sensors to be deployed, and with the initial values of β_1 and γ equal to 0.4 and 0.025 respectively. Referring to Equation (12), the values of β_2, β_3, β_4, β_5 and β_6 were respectively 0.35, 0.30, 0.25, 0.20 and 0.15.

For $\alpha = \beta_7$, considering Equation (9), this results in a LS-WSN of heterogeneity level equal to 6. This means that the network will be composed of 6 types of sensors. According to Equation (10), we obtain Equation (24):

$$\begin{aligned} ((\alpha - \beta_1) \times (1 + (\alpha - \beta_2) \times (1 + (\alpha - \beta_3) \\ \times (1 + (\alpha - \beta_4) \times (1 + (\alpha - \beta_5) \times (1 + (\alpha - \beta_6)))))))) = 1 \end{aligned} \tag{24}$$

We can easily determine the value of the first parameter of our model, knowing the respective values of β_i, where i being an integer from 1 to 7. Then calculating this value gives $\alpha = 0.8544$. For the different values of β_i and $\alpha = 0.8544$, we can count the different types of nodes using Equation (18). The results are enumerated hereinafter:

- type-1. $N \times (\alpha - \beta_1) = 10000 \times (0.8608 - 0.40) = 4608$
- type-2. $N \times (\alpha - \beta_1)(\alpha - \beta_2) = 6000 \times (0.8544 - 0.40)(0.8544 - 0.35) = 2354$
- type-3. $N \times (\alpha - \beta_1)(\alpha - \beta_2)(\alpha - \beta_3) = 6000 \times (0.8544 - 0.40)(0.8544 - 0.35)(0.8544 - 0.30) = 1320$
- type-4. $N \times (\alpha - \beta_1) \times (\alpha - \beta_2) \times (\alpha - \beta_3) \times (\alpha - \beta_4) = 6000 \times (0.8544 - 0.40) \times (0.8544 - 0.35) \times (0.8544 - 0.30) \times (0.8544 - 0.25) = 806$
- type-5. $N \times (\alpha - \beta_1) \times (\alpha - \beta_2) \times (\alpha - \beta_3) \times (\alpha - \beta_4) \times (\alpha - \beta_5) = 6000 \times (0.8544 - 0.40) \times (0.8544 - 0.35) \times (0.8544 - 0.30) \times (0.8544 - 0.25) \times (0.8544 - 0.20) = 534.$
- type-6. $N \times (\alpha - \beta_1) \times (\alpha - \beta_2) \times (\alpha - \beta_3) \times (\alpha - \beta_4) \times (\alpha - \beta_5) \times (\alpha - \beta_6) = 6000 \times (0.8544 - 0.40) \times (0.8544 - 0.35) \times (0.8544 - 0.30) \times (0.8544 - 0.25) \times (0.8544 - 0.20) = 378.$

Thus, the number of type-1, type-2, type-3, type-4, type-5, and type-6 of sensors for the different levels of heterogeneity are detailed in Table 2 and Figure 7. As for the categorization of the energy resources of the sensors, we set the amount of energy of sensors of type-1, noted E_1 and the value of the parameter γ at 0.2 J and 0.5 J respectively. By using Equation (9), we obtained the energy of the other types of sensors summarized in the Table 3, namely $E_2 = 0.4$ J, $E_3 = 0.5$ J, $E_4 = 0.6$ J, $E_5 = 7$ J and $E_5 = 8$ J for the LS-WSN with heterogeneity level equal to 6.
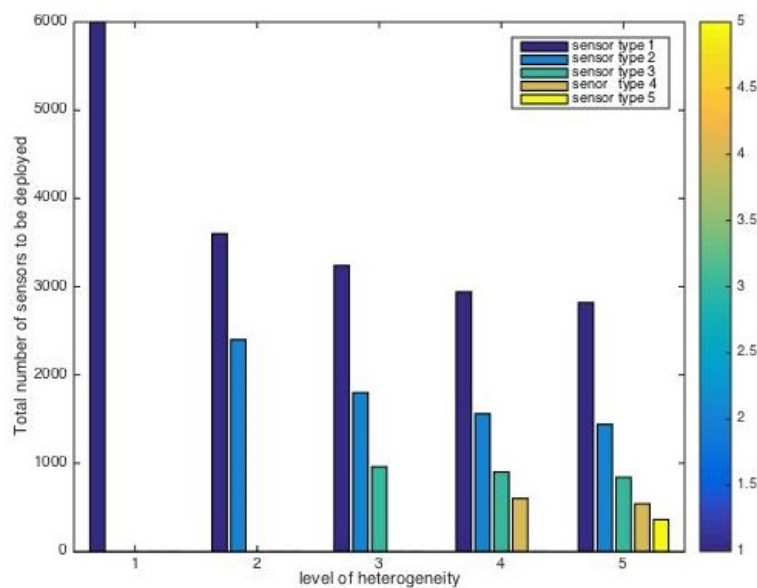
Figure 7. Distribution of a set of sensors according to various levels of heterogeneity.

Table 2. Deployment of 10,000 sensors in six heterogeneity level scenarios.

	LS-WSN-1	LS-WSN-2	LS-WSN-3	LS-WSN-4	LS-WSN-5	LS-WSN-6
Type-1	10,000	6000	5200	4900	4700	4608
Type-2	n/a	4000	3000	2600	2400	2354
Type-3	n/a	n/a	1800	1500	1400	1320
Type-4	n/a	n/a	n/a	1000	900	806
Type-5	n/a	n/a	n/a	n/a	600	533
Type-6	n/a	n/a	n/a	n/a	n/a	378

Table 3. Energy distribution by level of heterogeneity.

	LS-WSN1	LS-WSN2	LS-WSN3	LS-WSN4	LS-WSN5	LS-WSN6
Type-1	0.2 J	0.2 J	0.2 J	0.2 J	0.2 J	0.2 J
Type-2	n/a	0.4 J	0.4 J	0.4 J	0.4 J	0.4 J
Type-3	n/a	n/a	0.5 J	0.5 J	0.5 J	0.5 J
Type-4	n/a	n/a	n/a	0.6 J	0.6 J	0.6 J
Type-5	n/a	n/a	n/a	n/a	0.7 J	0.7 J
Type-6	n/a	n/a	n/a	n/a	n/a	0.8 J

Through this demonstration, we have proven that the determination of the different sensors required for a LS-WSN is possible. We can thus easily quantify for a given level of heterogeneity, the different number of sensors.

5.2. Evaluation of the Proposed Clustering Algorithm for LS-WSN

We evaluated our clustering algorithm in the following metrics: lifetime, energy consumption, the number of packets transmitted to the BS, and the effect of clustering in the energy consumption. We present the results of our algorithm by varying the scope of the nodes and then compare it to LEACH [32], E-LEACH [40], SEP [41], DEEC [37], Modified E-LEACH [42], EECDA [43], DSCHE [44], and BEENISH [45].

5.2.1. Lifetime

We study the lifetime of LS-WSNs according to the level of network heterogeneity. This lifetime is expressed in terms of the number of sensors still functional after an activity time expressed as the

number of data sent to the BS as shown in Figure 8. Moreover, Table 4 summarizes the number of rounds for which the first and last sensor dies of each level of heterogeneity draw their energies. For instance, for the case of LS-WSN-3, the first sensors exhaust their energy after 356 rounds, and the last ones after 2328 rounds, while for the LS-WSN-6, the first sensors exhaust their energy after 583 rounds, and the last ones after 3299 rounds.

Furthermore, functional sensors are sensors that have not yet exhausted their energies. In the homogeneous LS-WSNs scenarios, sensors exhaust their energy around the 668th data sending round, whereas for the heterogeneous LS-WSNs in all its variant, all sensors keep their energy as long as possible, as shown in Figure 8. In the case of LS-WSN-2, the first sensors exhaust their energy at the 269th round and the last ones exhaust at the 2158th round. In the case of LS-WSN-6, the first sensors exhaust their energy at the 583rd round and the last nodes die at the 3299th round. As it can be seen from Figure 8, among all levels of heterogeneity, the LS-WSN-6 offers the longest network lifetime. This allows us to conclude, in the context of the LS-WSNs, that sensor heterogeneity maximizes network lifetime.
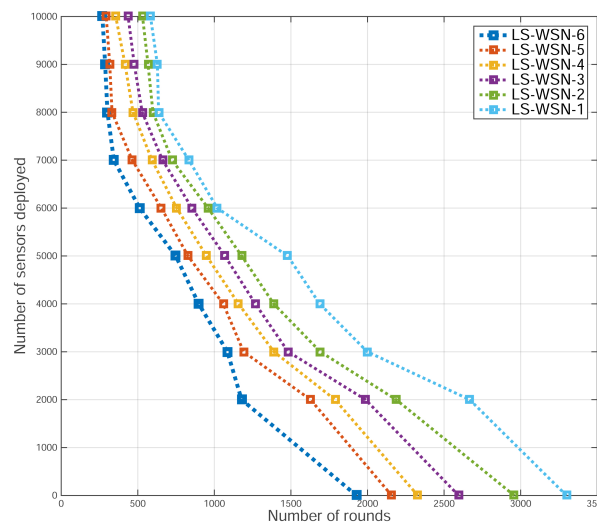

Figure 8. Lifetime of the network in terms of the number of active sensors vs. the number of rounds.

Table 4. Evolution of sensors energy depletion per level of heterogeneity.

	Rounds	
	First Sensor Dies	**Last Sensor Dies**
LS-WSN-1	269	1928
LS-WSN-2	294	2158
LS-WSN-3	356	2328
LS-WSN-4	437	2595
LS-WSN-5	531	2558
LS-WSN-6	583	3299

5.2.2. Throughput

We evaluated the amount of data transmitted to the BS over a period of time on the different levels of LS-WSN heterogeneity, as shown in Figure 9. This measurement refers to the amount of information collected by the network from the sensor and sent to the BS. The LS-WSN-6 sends the maximum amount of data to the BS of all variants as shown in Figure 9. The number of packets transferred to the BS by the LS-WSN-1, LS-WSN-2, LS-WSN-3, LS-WSN-4, LS-WSN-5, and LS-WSN-6 are respectively, 62.0×10^4, 0.90×10^4, 1.16×10^4, 1.71×10^4, 2.20×10^4, and 2.79×10^4.
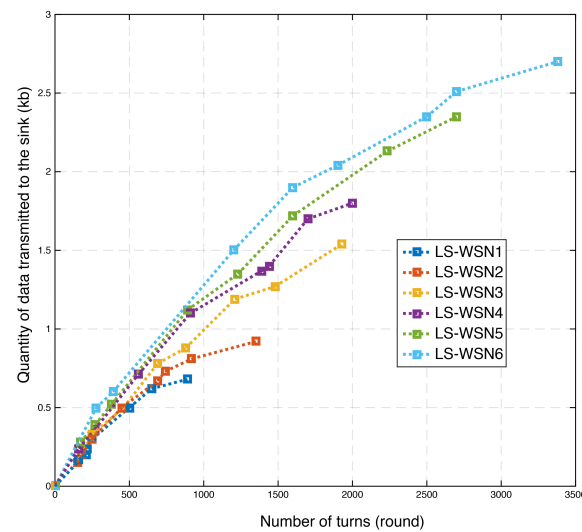
Figure 9. Quantity of data sent by from the considered LS-WSNs to the base station (BS).

5.2.3. Power Consumption

To confirm the hypothesis of sensor heterogeneity in the LS-WSN life extension, we evaluated for each network the sum of energy dissipated in a given period and according to its level of heterogeneity, as shown in Figure 10. This measurement refers to the instantaneous amount of energy exhausted by a network during a data transfer cycle, i.e., the difference in energy from the beginning of the cycle to its end. Here, the total initial energies were 20.0 J, 28.0 J, 31.0 J, 34.0 J, 36.0 J and 38.0 J, for *type*-1, *type*-2, *type*-3, *type*-4, *type*-5, and *type*-6 sensors respectively. As it can be seen from the results presented in Figure 10, the LS-WSN-6 outperformed the LS-WSN-1, LS-WSN-2, LS-WSN-3, LS-WSN-4 and LS-WSN-5. Thus, the energy dissipation (in Joules) rate was a decreasing function of the increasing level of heterogeneity of the LS-WSN.
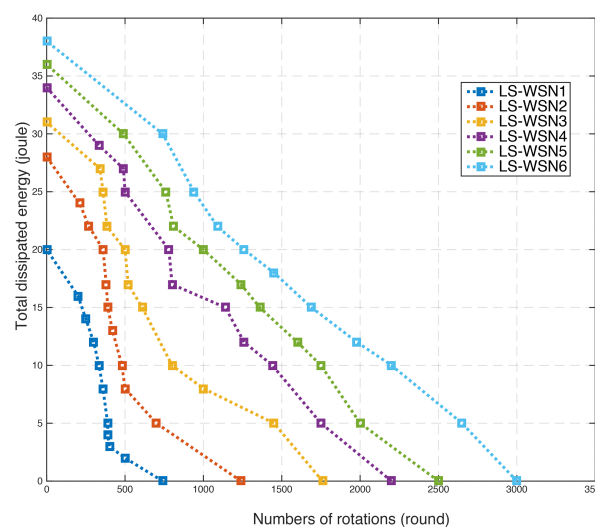

Figure 10. Total energy dissipated.

5.2.4. Effect of Clustering in the Energy Consumption

Besides, the LS-WSN architecture we have proposed in this paper is a three-tier architecture based on clustering. We sought to identify the impact of the range of the different sensors on the formation of clusters, the energy consumption of clusters in the different levels of heterogeneity of the networks.

From results presented in Figure 11, we noticed that the percentage of the consumed energy remained low (about 0.088%) in the clusters and did not exceed 0.146% in the worst case when the average range oscillated between 10 and 20 meters. These values remained reasonable for a network composed of 10, 000 heterogeneous sensors.
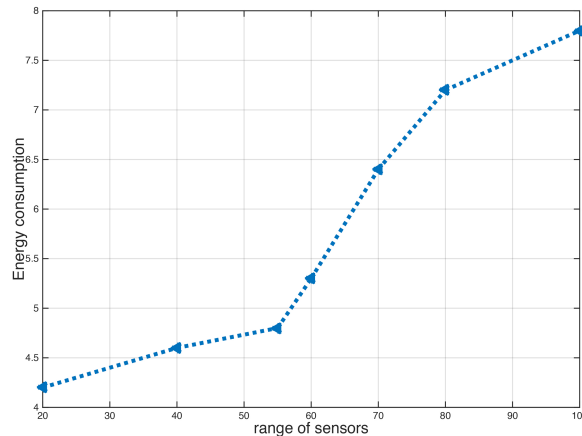

Figure 11. Energy consumption of the sensors vs. the range of sensors.

On the other hand, Figure 12 shows that the number of CHs built in the different heterogeneous networks depends on the range of sensors. We noticed that the number of CHs decreases regularly as a function of the range of the nodes. This is explained by the fact that the increase of the range leads to the increase of the number of neighbors (degree) for each node. Under these conditions, the number of members in a cluster increases and consequently the number of CHs created decreases.
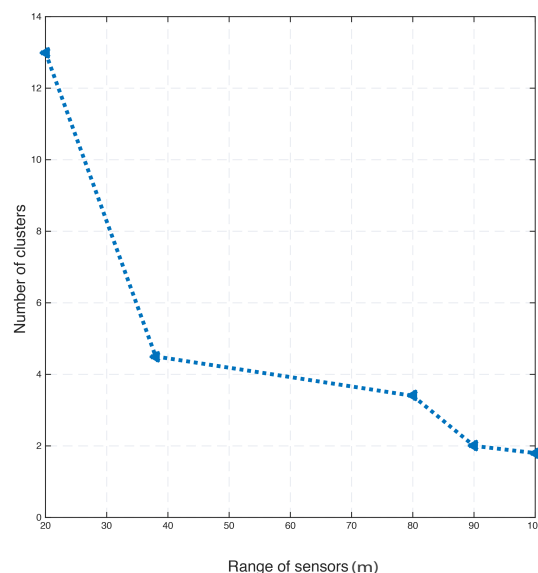

Figure 12. Number of clusters formed vs. the range of sensors.

From results presented in Figure 13, it can be observed that the evolution of the average number of packets sent to the BS (resp. to the CHs) was a function of the range of the nodes. The number of packets sent to the CH increased steadily with the range of the sensors, this is due to the amplification of the cluster members by multiplying their coverage areas. Moreover, Figure 13 shows that the second

curve was the opposite of the first. Indeed, increasing the range of the nodes minimized the number of created CHs and subsequently reduced the number of packets sent to the BS.
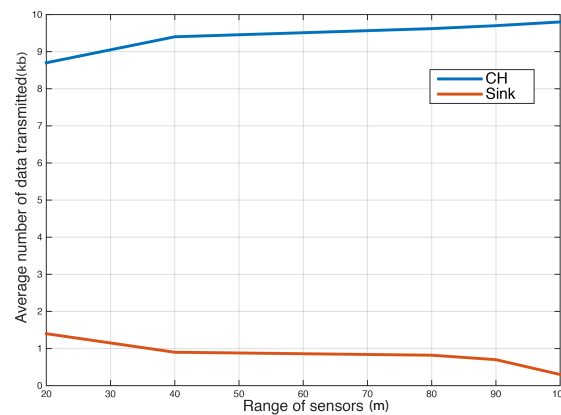

Figure 13. Mean number of transmitted data according to the range of sensors.

5.3. Performance Comparison

Some performance of protocols dedicated to heterogeneous LS-WSNs have been computed and compared to our approach of using our scenarios of heterogeneous LS-WSNs for big data collection. As shown in the Table 5, considering the same number of sensors to deploy and the same initial energy quantity of the sensors, we noticed that our different models presented a set of performances better than the compared protocols dedicated to heterogeneous LS-WSN.

Table 5. Comparison of our proposed LS-WSN-1, LS-WSN-2, and LS-WSN-3 vs. some existing protocols in terms of total number of rounds.

	No. of Sensors	Total Energy	Level of Heterogeneity	No. of Rounds
LEACH [32]	10,000	68 J	1	678
E-LEACH [40]	10,000	68 J	1	893
SEP [41], DEEC [37]	10,000	68 J	1	1348
Modified E LEACH [42]	10,000	68 J	1	1542
EECDA [43]	10,000	68 J	2	1621
DSCHE [44]	10,000	68 J	2	1968
BEENISH [45]	10,000	68 J	3	2159
LS-WSN-1	10,000	68 J	1	1878
LS-WSN-2	10,000	68 J	2	2926
LS-WSN-3	10,000	68 J	3	3675

Furthermore, Figure 14 represents the average number of clusters constructed as a function of the cardinality of the network for the different algorithms. We noticed as well that the number of clusters increased regularly with the size of the network. Moreover, from Figure 14, we noticed that our algorithm produced fewer clusters in most cases (size between 40 and 200) compared to protocols such as LEACH and DSCHE whose number of clusters increased while scaling of the number sensors. Finally, the specificity of our algorithm lay in the fact that the same number of CHs could be used to manage a network of increasing size. This is explained by the efficiency of the structure created by our algorithm in its ability to adapt when adding sensors to the network.

Subsequently, we calculated the amount of energy consumed by the protocols that structure the LS-WSN according to a clustering-based architecture. Figure 15 represents the average energy consumption of the different clusters in these protocols. From plotted results, we can see that the values obtained by our approach were quite low (almost half) compared to those obtained by most of the

studied protocols. These results show that our clustering algorithm minimized energy consumption, effectively extended the network lifetime and ensured a good big data collection process.
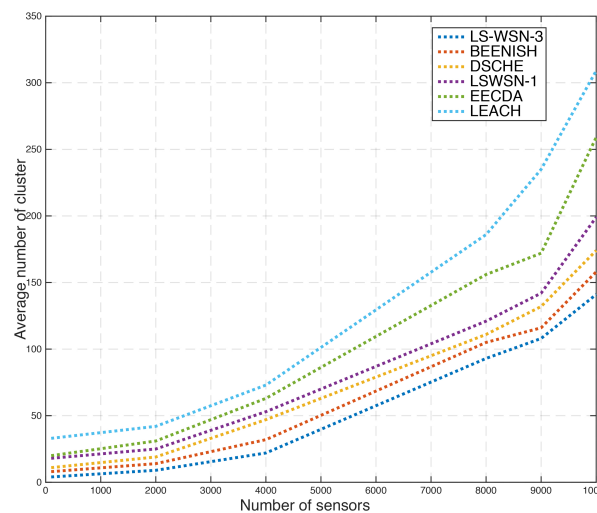

Figure 14. Comparison of the average number of clusters according to the number of sensors.
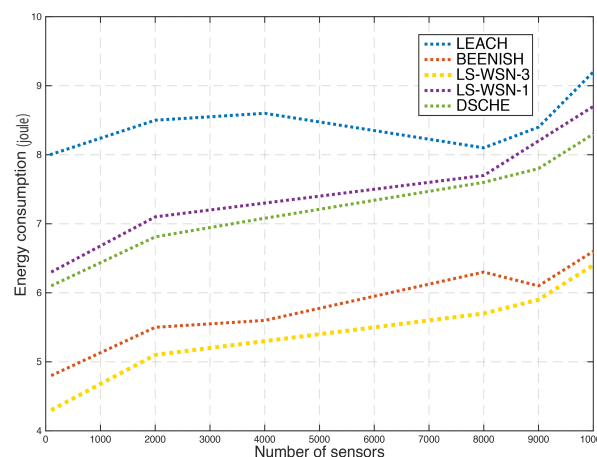

Figure 15. Comparison of the energy consumed by the compared protocols.

6. Conclusions

There has been a lot of interest concerning LS-WSNs issues in recent years. On the one hand, the connectivity and better coverage of these kinds of networks remains a challenge for monitoring and data collection applications. On the other hand, the issue of energy consumption remains very critical as the number of sensors becomes more and more important, which requires the development of new sensor energy conservation techniques to extend the life of these networks. In this paper, we are interested in the use of network types for the collection of big data in smart cities. Given the intersecting characteristics of big data, we have chosen to use heterogeneous LS-WSNs. To build this network, we proposed a mathematical quantification model, which, starting from a given level of heterogeneity, determines the number of different sensors and the energy associated with the latter. Then, we proposed a three-third network deployment architecture. To do this, we proposed an algorithm that allows the sensors to be organized in clusters following a three-tier architecture. For the experimental validation of the model, we considered LS-WSNs with up to six levels of heterogeneity, i.e., networks made up of at least six different types of sensors dedicated to the collection of big data from smart cities. Our experiments have shown that the proposed model can describe any level of heterogeneity of LS-WSN.

As for the proposed clustering algorithm, it significantly extends the lifetime of the network as the level of network heterogeneity increases. In the case of the 6th heterogeneous network, the network lifetime increases by 348.65% for the 84% increase in network energy. The comparison of our clustering approach with some well-known protocols in the literature shows that our algorithm brings a gain in terms of improved topology and energy conservation at the sensor level. Finally, we have shown that the sensor range influences the clustering and the energy consumption of the sensors. The level of sensor heterogeneity in LS-WSNs significantly reduces the aggregation delay and improves the latency in the data collection process in LS-WSNs. In our future works, we envision contributing to a deterministic fault-tolerant deployment strategy that aims at adjusting the topology of LS-WSNs to correct connectivity and coverage gaps, as well as extending the life of the network.

Author Contributions: Conceptualization, A.C.D., A.A.A., A.M.G. and A.M.; formal analysis, A.C.D., A.A.A., A.M.G. and O.T.; funding acquisition, A.M.G. and Z.A.; investigation, A.C.D., A.A.A., A.M.G. and O.T.; Methodology, A.C.D., A.A.A., A.M.G. and A.M.; project administration, A.A.A. and A.M.G.; supervision, A.A.A., A.M.G., A.M. and Z.A.; Validation, A.C.D., A.A.A. and O.T.; writing—original draft, A.C.D. and A.A.A.; writing—review and editing, A.A.A., A.M.G., A.M., O.T. and Z.A. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Djedouboum, A.; Ari, A.A.A.; Gueroui, A.; Mohamadou, A.; Aliouat, Z. Big Data Collection in Large-Scale Wireless Sensor Networks. *Sensors* **2018**, *18*, 4474. [CrossRef] [PubMed]
2. Ari, A.A.A.; Gueroui, A.; Titouna, C.; Thiare, O.; Aliouat, Z. Resource allocation scheme for 5G C-RAN: A Swarm Intelligence based approach. *Comput. Netw.* **2019**, *165*, 106957. [CrossRef]
3. Martínez-de Dios, J.R.; de San Bernabé, A.; Viguria, A.; Torres-González, A.; Ollero, A. Combining unmanned aerial systems and sensor networks for earth observation. *Remote Sens.* **2017**, *9*, 336. [CrossRef]
4. Gbadouissa, J.E.Z.; Ari, A.A.A.; Titouna, C.; Gueroui, A.M.; Thiare, O. HGC: HyperGraph based Clustering scheme for power aware wireless sensor networks. *Future Gener. Comput. Syst.* **2020**, *105*, 175–183. [CrossRef]
5. Baker, T.; Aldawsari, B.; Asim, M.; Tawfik, H.; Maamar, Z.; Buyya, R. Cloud-SEnergy: A bin-packing based multi-cloud service broker for energy efficient composition and execution of data-intensive applications. *Sustain. Comput. Inform. Syst.* **2018**, *19*, 242–252. [CrossRef]
6. Statista Research Department. *Internet of Things (IoT) Connected Devices Installed Base Worldwide from 2015 to 2025 (in Billions)*; Statista Research Department, Ss. Cyril and Methodius University in Skopje, Faculty of Computer Science and Engineering: Skopje, North Macedonia, 2020.
7. Nzegha, A.F.; Fendji, J.L.E.; Thron, C.; Tayou, C.D. Improving Deep Unconstrained Facial Recognition by Data Augmentation. In *Implementations and Applications of Machine Learning*; Springer: Cham, Switzerland, 2020; pp. 179–195.
8. Ari, A.A.A.; Gueroui, A.; Labraoui, N.; Yenke, B.O. Concepts and evolution of research in the field of wireless sensor networks. *Int. J. Comput. Netw. Commun.* **2015**, *7*, 81–98. [CrossRef]
9. Kim, B.S.; Kim, K.I.; Shah, B.; Chow, F.; Kim, K.H. Wireless sensor networks for big data systems. *Sensors* **2019**, *19*, 1565. [CrossRef]
10. Aboubakar, M.; Kellil, M.; Bouabdallah, A.; Roux, P. Using Machine Learning to Estimate the Optimal Transmission Range for RPL Networks. In Proceedings of the NOMS 2020-2020 IEEE/IFIP Network Operations and Management Symposium, Budapest, Hungary, 20–24 April 2020; IEEE: New York, NY, USA, 2020; pp. 1–5.
11. Marrero, D.; Suárez, A.; Macías, E.; Mena, V. Extending the Battery Life of the ZigBee Routers and Coordinator by Modifying Their Mode of Operation. *Sensors* **2020**, *20*, 30. [CrossRef]
12. Gaboitaolelwe, J.; Zungeru, A.M.; Chuma, J.; Ditshego, N.; Semong, T. A Formal Analytical Modeling and Simulation of Wireless Sensor Home Network. *Int. J. Intell. Eng. Syst.* **2020**, *13*, 56–68.

13. Ari, A.A.A.; Damakoa, I.; Gueroui, A.; Titouna, C.; Labraoui, N.; Kaladzavi, G.; Yenké, B.O. Bacterial foraging optimization scheme for mobile sensing in wireless sensor networks. *Int. J. Wirel. Inf. Netw.* **2017**, *24*, 254–267. [CrossRef]

14. Manfreda, S.; McCabe, M.F.; Miller, P.E.; Lucas, R.; Pajuelo Madrigal, V.; Mallinis, G.; Ben Dor, E.; Helman, D.; Estes, L.; Ciraolo, G.; et al. On the use of unmanned aerial systems for environmental monitoring. *Remote Sens.* **2018**, *10*, 641. [CrossRef]

15. Njoya, A.N.; Ari, A.A.A.; Awa, M.N.; Titouna, C.; Labraoui, N.; Effa, J.Y.; Abdou, W.; Gueroui, A. Hybrid Wireless Sensors Deployment Scheme with Connectivity and Coverage Maintaining in Wireless Sensor Networks. *Wirel. Pers. Commun.* **2020**, *112*, 1893–1917. [CrossRef]

16. Sambo, D.W.; Forster, A.; Yenke, B.O.; Sarr, I.; Gueye, B.; Dayang, P. Wireless Underground Sensor Networks Path Loss Model for Precision Agriculture (WUSN-PLM). *IEEE Sens. J.* **2020**, *20*, 5298–5313. [CrossRef]

17. Mészáros, L.; Varga, A.; Kirsche, M. Inet framework. In *Recent Advances in Network Simulation*; Springer: Cham, Switzerland, 2019; pp. 55–106.

18. Wu, X.; Zhu, X.; Wu, G.Q.; Ding, W. Data mining with big data. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 97–107.

19. Zikopoulos, P.; Eaton, C.; Zikopoulos, P. *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*; McGraw-Hill Osborne Media: New York, NY, USA, 2011.

20. Bhadani, A.K.; Jothimani, D. Big data: Challenges, opportunities, and realities. In *Effective Big Data Management and Opportunities for Implementation*; IGI Global: Hershey, PA, USA, 2016; pp. 1–24.

21. Chen, M.; Mao, S.; Liu, Y. Big data: A survey. *Mob. Netw. Appl.* **2014**, *19*, 171–209. [CrossRef]

22. Harb, H.; Idrees, A.K.; Jaber, A.; Makhoul, A.; Zahwe, O.; Taam, M.A. Wireless sensor networks: A big data source in Internet of Things. *Int. J. Sens. Wirel. Commun. Control* **2017**, *7*, 93–109. [CrossRef]

23. Vijayakumari, R.; Kirankumar, R.; Rao, K.G. Comparative analysis of google file system and hadoop distributed file system. *Int. J. Adv. Trends Comput. Sci. Eng.* **2014**, *3*, 553–558.

24. Farrah, S.; El Manssouri, H.; Ziyati, E.; Ouzzif, M. An approach to analyze large scale wireless sensors network data. *Int. Res. J. Comput. Sci. (IRJCS)* **2015**, *2*, 7–12.

25. Capriolo, E.; Wampler, D.; Rutherglen, J. *Programming Hive: Data Warehouse and Query Language for Hadoop*; O'Reilly Media, Inc.: Sevastopol, CA, USA, 2012.

26. Rios, L.G.; Diguez, J.A.I. Big data infrastructure for analyzing data generated by wireless sensor networks. In Proceedings of the 2014 IEEE International Congress on Big Data, Anchorage, AK, USA, 27 June–2 July 2014; IEEE: New York, NY, USA, 2014; pp. 816–823.

27. Hamidouche, R.; Aliouat, Z.; Ari, A.A.A.; Gueroui, M. An efficient clustering strategy avoiding buffer overflow in IoT sensors: A bio-inspired based approach. *IEEE Access* **2019**, *7*, 156733–156751. [CrossRef]

28. Kone, C.T. Conception de l'architecture d'un réseau de capteurs sans fil de Grande Dimension. Ph.D. Thesis, Université Henri Poincaré-Nancy I, Lorraine, France, 2011.

29. Hamidouche, R.; Khentout, M.; Aliouat, Z.; Gueroui, A.M.; Ari, A.A.A. Sink Mobility Based on Bacterial Foraging Optimization Algorithm. In *Computational Intelligence and Its Applications. CIIA 2018. IFIP Advances in Information and Communication Technology*; Amine, A., Mouhoub, M., Ait Mohamed, O., Djebbar, B., Eds.; Springer: Cham, Switzerland, 2018; Volume 522, p. 352.

30. Ari, A.A.A.; Yenke, B.O.; Labraoui, N.; Damakoa, I.; Gueroui, A. A power efficient cluster-based routing algorithm for wireless sensor networks: Honeybees swarm intelligence based approach. *J. Netw. Comput. Appl.* **2016**, *69*, 77–97. [CrossRef]

31. Ari, A.A.A.; Labraoui, N.; Yenke, B.O.; Gueroui, A. Clustering algorithm for wireless sensor networks: The honeybee swarms nest-sites selection process based approach. *Int. J. Sens. Netw.* **2018**, *27*, 1–13. [CrossRef]

32. Heinzelman, W.R.; Chandrakasan, A.; Balakrishnan, H. Energy-efficient communication protocol for wireless microsensor networks. In Proceedings of the 33rd Annual Hawaii Internationalconference on System Sciences, Maui, HI, USA, 7 January 2000; IEEE: New York, NY, USA, 2000; p. 10.

33. Farooq-i Azam, M.; Ayyaz, M.N. Location and position estimation in wireless sensor networks. In *Wireless Sensor Networks: Current Status and Future Trends*; CRC Press, Taylor & Francis Group: Boca Raton, FL, USA, 2012; pp. 179–214.

34. Instruments, T. CC2420: 2.4 GHz IEEE 802.15. 4/ZigBee-Ready RF Transceiver. 2007. Available online: http://www.ti.com/lit/ds/symlink/cc2420.pdf (accessed on January 12 2020).

35. Cai, X.; Duan, Y.; He, Y.; Yang, J.; Li, C. Bee-sensor-C: An energy-efficient and scalable multipath routing protocol for wireless sensor networks. *Int. J. Distrib. Sens. Netw.* **2015**, *11*, 976127. [CrossRef]

36. Eslami, F.; Sima, M. Capacitive boosting for fpga interconnection networks. In Proceedings of the 2011 21st International Conference on Field Programmable Logic and Applications, Chania, Greece, 5–7 September 2011; IEEE: New York, NY, USA, 2011; pp. 453–458.

37. Qing, L.; Zhu, Q.; Wang, M. Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks. *Comput. Commun.* **2006**, *29*, 2230–2237. [CrossRef]

38. Singh, S. Energy efficient multilevel network model for heterogeneous WSNs. *Eng. Sci. Technol. Int. J.* **2017**, *20*, 105–115. [CrossRef]

39. Rani, S.; Ahmed, S.H.; Talwar, R.; Malhotra, J. Can sensors collect big data? An energy-efficient big data gathering algorithm for a WSN. *IEEE Trans. Ind. Inform.* **2017**, *13*, 1961–1968. [CrossRef]

40. Heinzelman, W.B.; Chandrakasan, A.P.; Balakrishnan, H. An application-specific protocol architecture for wireless microsensor networks. *IEEE Trans. Wirel. Commun.* **2002**, *1*, 660–670. [CrossRef]

41. Smaragdakis, G.; Matta, I.; Bestavros, A. *SEP: A Stable Election Protocol for Clustered Heterogeneous Wireless Sensor Networks*; Technical Report; Boston University Computer Science Department: Silber Way, Boston, MA, USA, 2004.

42. Sedighimanesh, A.; Sedighimanesh, M.; Baqeri, J. Improving wireless sensor network lifetime using layering in hierarchical routing. In Proceedings of the 2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI), Tehran, Iran, 5–6 November 2015; IEEE: New York, NY, USA, 2015; pp. 1145–1149.

43. Kumar, D.; Aseri, T.C.; Patel, R. EECDA: Energy efficient clustering and data aggregation protocol for heterogeneous wireless sensor networks. *Int. J. Comput. Commun. Control* **2011**, *6*, 113–124. [CrossRef]

44. Kumar, D. Distributed stable cluster head election (DSCHE) protocol for heterogeneous wireless sensor networks. *Int. J. Inf. Technol. Commun. Converg.* **2012**, *2*, 90–103. [CrossRef]

45. Qureshi, T.; Javaid, N.; Khan, A.; Iqbal, A.; Akhtar, E.; Ishfaq, M. BEENISH: Balanced energy efficient network integrated super heterogeneous protocol for wireless sensor networks. *Procedia Comput. Sci.* **2013**, *19*, 920–925. [CrossRef]