Abstract: A wireless sensor network (WSN) is a modern technology in radio communication. A WSN comprises a number of sensors that are randomly spread in a specific area for sensing and monitoring physical attributes that are difficult to monitor by humans, such as temperature, humidity, and pressure. Many problems, including data routing, power consumption, clustering, and selecting cluster heads (CHs), may occur due to the nature of WSNs. Various protocols have been conducted to resolve these issues. One of the important hierarchical protocols that are used to reduce power consumption in WSNs is low-energy adaptive clustering hierarchy (LEACH). This paper presents a comprehensive study of clustering protocols for WSNs that are relevant to LEACH. This paper is the first to emphasize on cluster formation and CHs selection methods and their strengths and weaknesses. A new taxonomy is presented to discuss LEACH variants on the basis of different classes, and the current survey is compared with other existing surveys. A complete comparison of the location, energy, complexity, reliability, multi-hop path, and load balancing characteristics of LEACH variants is conducted. Future research guidelines for CHs selection and cluster formation in WSNs are also discussed.

Keywords: cluster formation; LEACH protocol; energy consumption; wireless sensor networks

1. Introduction

Research interest in wireless sensor networks (WSNs) has increased significantly during last few decades [1] because of the increased use of WSNs for many purposes, such as monitoring environments to collect data on physical attributes (e.g., temperature and pressure), road traffic monitoring, health monitoring, and transportation [2–8]. WSNs comprise a huge number of limited-power sensors that sense and send data to base stations for further processing [9].

Power consumption is a challenge in WSNs. Sensors are usually positioned randomly in a large zone; if one of the sensors depletes its energy, replacing the power supply becomes difficult, especially in hostile inaccessible environments, and energy holes emerge [10]. Furthermore, the power of a sensor node decreases during the transmission of data [11,12]. Therefore, routing is important in WSNs, and proper selection of effective routing protocols can prolong network lifetime [13]. Many protocols have been presented to minimize the power consumption in WSNs [14,15]. These protocols depend on several factors, such as network structure, protocol operations, and path establishment [16]. Designing and developing routing protocols are challenging to researchers because of the limited power of sensors in WSNs.
Owing to the difficulty and inefficiency of replacing the power supply of dead nodes (because the number of nodes can be large and deployment can be arbitrary), meeting application requirements becomes challenging [13,16].

This survey discusses routing protocols that have been conducted to deal with these challenges to evaluate the efficiency of these protocols and determine their drawbacks.

One of the network structure categories is clustering or hierarchical routing protocols [17]. Sensors are self-organized, and clusters are usually formed when the deployment completes in clustering routing protocols. Each cluster contains a private node called the cluster head (CH) that performs extra tasks, such as collecting data from normal nodes in a cluster, compressing the data, and sending them to the base station (BS) for analysis and decision-making [18–21]. These additional tasks cause the CH to expend the extra amount of energy than other normal nodes do in the cluster.

One of the early clustering routing protocols is the low-energy adaptive clustering hierarchy (LEACH), which is proposed by Heinzelmann [19] to increase the network lifetime by rotating the CH’s tasks among the nodes. However, this protocol has limitations in energy distribution in clusters and in maximizing network lifetime. Thus, many studies have enhanced the LEACH protocol to balance the energy distribution among sensors in clusters [22,23] and increase network lifetime [24,25].

LEACH possesses several drawbacks [26,27], such as CHs being selected randomly without taking into account the residual energy of a node or the distance from a node to the BS [28]. Another problem is the transmission of data from a CH to the BS because the communication is conducted using the single-hop method. This method causes a serious issue when the CH is far from the BS because extra amount of energy is used than when the CH is close to the BS, resulting in energy holes; in which isolated nodes cannot send their data. This situation considerably affects network performance. Therefore, the development of efficient clustering routing protocols is a significant part of research on WSNs to increase network lifetime [29].

Many surveys [30–33] have examined LEACH enhancement clustering protocols by using different classifications based on different criteria, such as the way of communication in routing protocols, node mobility, and node location awareness. This survey presents an organized and comprehensive review of LEACH enhancement clustering routing protocols by using a new taxonomy depends on different parameters, such as distance, routing path, and network coverage, with emphasis on clustering technology and challenges.

The classification in this survey depends on CH selection and cluster formation methods, which are new classifications of LEACH enhancement protocols for WSNs in term of distance, energy efficiency, residual energy, and network coverage. The rationale behind focusing on the CH selection and the cluster formation categories in this survey is due to their importance in the hierarchical-based protocols because the CH selection and the cluster formation play a significant role in prolonging the network lifetime. The difference between these categories is that the CHs are firstly selected based on some criteria such as the energy and the distance to the BS. After that, the clusters will be formed by joining each normal node to its suitable CH based on different methods. This classification is expected to provide researchers a clear view of such protocols and guide them in proposing new efficient cluster formation and CH selection methods.

Recent LEACH enhancement clustering routing protocols for WSNs will be discussed based on CH selection and cluster formation methods. Each clustering routing protocol is examined in detail, and each CH selection method and its specifications and obstacles are emphasized. Summarization and comparisons of clustering routing protocols for WSNs are performed based on the parameters that are used to select CHs and the parameters that are used to form clusters.

The comparison of different protocols has been done using different performance elements, like distance, reliability, and complexity. The last part of the paper presents several important matters that are necessary to be considered when proposing new clustering routing protocols or CH selection methods for WSNs.

The contributions of this survey are as follows:
Hierarchical routing protocols based on LEACH are classified based on cluster formation and CH selection methods by using a new taxonomy based on new parameters, such as residual energy, distance, and network coverage.

A large number of LEACH–based protocols based on the new taxonomy are discussed.

A comprehensive survey is performed to discuss LEACH enhancements in detail.

Qualitative comparisons with other existing surveys are conducted.

A comparison among protocols is performed in terms of several parameters (e.g., energy, overhead, and complexity) to help researchers select the most efficient protocol based on their needs.

This survey is organized as follows: Section 2 provides a background of WSNs and an overview of the LEACH protocol. In addition, LEACH-based routing techniques and their classification into two main categories, namely, CH selection-based and cluster formation-based protocols are analyzed and discussed. Routing protocols that belong to each category are also discussed in detail. In Section 3, a comparative analysis of different protocols is described in detail. The future issues in designing routing protocols are discussed in Section 4. The survey is presented in Section 5.

2. Background and Overview of WSNs

Materials and Methods should be described with sufficient details to allow others to replicate and build on published results. Please note that publication of your manuscript implicates that you must make all materials, data, computer code, and protocols associated with the publication available to readers. Please disclose at the submission stage any restrictions on the availability of materials or information. New methods and protocols should be described in detail while well–established methods can be briefly described and appropriately cited.

Research manuscripts reporting large datasets that are deposited in a publicly available database should specify where the data have been deposited and provide the relevant accession numbers. If the accession numbers have not yet been obtained at the time of submission, please state that they will be provided during review. They must be provided prior to publication.

Interventionary studies involving animals or humans, and other studies require ethical approval must list the authority that provided approval and the corresponding ethical approval code.

This section presents an overview of WSNs, and their characteristics, challenges, and components are discussed. An overview of the LEACH protocol is also presented.

2.1. WSNs

WSNs are sets of sensors deployed for sensing and monitoring an environment to collect data on natural attributes, such as temperature and pressure [34,35]. The data collected by each sensor are transmitted to a BS, in which analysis of the collected data is conducted [36]. WSNs are useful to monitor and collect data in areas where humans cannot do so.

These areas include healthcare (monitoring patients and related medical management), providing emergency services in cases of disasters (e.g., earthquakes), and traffic control [37,38]. WSNs differ from traditional networks because the former consists of many small nodes called sensors, but these sensors have many limitations [39]. Thus, WSNs have distinctive characteristics, which are listed in Table 1.
Table 1. WSN’s characteristics.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computing capabilities</td>
<td>Sensor nodes in WSNs are small in size and memory and have a limited battery power [40].</td>
</tr>
<tr>
<td>Battery energy</td>
<td>Sensor nodes consume energy to operate functions, such as sensing, gathering and transmitting data. Power consumption is one of the challenges in WSNs. Given that the sensors are usually deployed randomly in a large area, if one of these nodes depletes its energy, replacing the power supply is difficult, especially in hostile and inaccessible environments [41].</td>
</tr>
<tr>
<td>Communication capabilities</td>
<td>Sensor nodes are affected by their surroundings, including rain, storms, mountains, and buildings. The communication distance is short, and the communication bandwidth is narrow [42].</td>
</tr>
<tr>
<td>Dynamic</td>
<td>In WSNs, several sensor nodes may be added or removed from the network according to task requirements. A sensor node may exit from the network when its energy is exhausted or because of other failures, thereby changing the topology of the network [43].</td>
</tr>
<tr>
<td>No center, self–organization</td>
<td>In WSNs, installing network infrastructure is unnecessary because a sensor node can adapt itself, and an independent network can automatically be formed after nodes switch on [44].</td>
</tr>
<tr>
<td>Multi–hop communications</td>
<td>In WSNs, a sensor node directly communicates with its neighbors, but if a node needs to communicate with other nodes, this communication must be via a multi–hop route through intermediate nodes. In contrast to traditional wired networks that need gateways and routers for multi–hop communication, nodes in WSNs act as routers in addition to their main functions as data collectors and senders [45].</td>
</tr>
<tr>
<td>Application relevance</td>
<td>WSNs are different from traditional networks in application relevance. WSNs are highly dependent on applications; thus, the topology, routing protocols, and physical signals in one network cannot be applied to others [46].</td>
</tr>
</tbody>
</table>

Challenges are also present in the WSN field [47,48], and researchers continue to develop new methods and strategies to address these challenges. Several of these challenges are summarized in Table 2.

Table 2. WSN’s challenges.

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limited battery power</td>
<td>One of the most significant obstacles in WSNs is power limitation because each sensor node has a limited power supply. Developing methods to maximize network lifetime by saving energy and balancing energy consumption remains a challenge [7,49].</td>
</tr>
<tr>
<td>Limited computation capability</td>
<td>Owing to the small processors, memory modules, and computational capabilities in sensor nodes, designing efficient algorithms that minimize complexities is a challenge [6].</td>
</tr>
<tr>
<td>Limited communication capability</td>
<td>The communication range in WSNs is roughly several hundred meters. Thus, developing algorithms to complete the transmission, analysis, and mining of sensed data with minimal communication complexity is important in WSNs [10].</td>
</tr>
</tbody>
</table>
Table 2. Cont.

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-hop communication</td>
<td>In WSNs, sensor nodes transmit data only to their neighbors, and this situation becomes a problem when a node must send data to other nodes. Thus, developing a routing algorithm to select a transmission route is a significant issue [50,51].</td>
</tr>
<tr>
<td>Nodes mobility</td>
<td>In mobile sensor networks, sensor nodes move frequently. Reconstructing routes to reduce communication failures and make routes adaptive remains a challenge in WSNs [52,53].</td>
</tr>
<tr>
<td>Large number of nodes</td>
<td>WSNs, that are deployed in a large area, comprise a large number of sensor nodes. Thus, considerations when designing algorithms include making sensor networks fault-tolerant and maintenance of these networks [54,55].</td>
</tr>
<tr>
<td>Sensor deployment</td>
<td>WSNs are used to monitor a large region by dropping a large number of nodes from an aircraft. How can these nodes be deployed efficiently to cover the entire region? [56,57]</td>
</tr>
<tr>
<td>Topology control</td>
<td>The power of nodes in WSNs is limited, and after a period, several nodes die. Changing the topology of the network, especially in dynamic WSNs, is an important issue [58].</td>
</tr>
</tbody>
</table>

2.2. WSN’s Components

A WSN is used to detect or monitor an area to collect data for further processing. Monitoring is conducted by a set of nodes, which are spread randomly in that area. Each node senses specific data and sends them to another node (if the communication model is multi-hop) until these data reach the BS, where data processing is performed on the data. Data are sent from one node to another node via radio transmission [59,60]. Figure 1 illustrates the components of a common WSN.

![Figure 1. WSN’s components.](image)

Sensor nodes are limited in size and weight, and each node in a WSN consists of a power supply (e.g., a battery), a small memory module, a small CPU, and an antenna, as shown in Figure 2.

Limited power is the main drawback of WSNs because sensors can consume their total energy in a short period due to sensing and transmission operations. The difficulty and inefficiency of replacing batteries or power supplies of dead nodes (because the number of nodes can be large and deployment can be arbitrary) pose serious challenges in meeting application requirements [61,62].
2.3. LEACH-Based Routing Protocols in WSNs

LEACH-based routing protocols are reviewed in this subsection. First, an overview of the LEACH protocol is provided. Second, several existing surveys are discussed. Finally, a new taxonomy of LEACH-based routing protocols in WSNs is presented based on the proposed categories, with focus on the advantages and disadvantages of each protocol.

2.3.1. Overview of the LEACH Protocol

The idea of the LEACH protocol [19] is to improve the efficiency of WSNs by reducing power consumption. A rotation technique of CH selection using a random factor has been proposed to achieve this goal.

The LEACH protocol operates in many rounds [63,64], and each of them comprises two stages, which are the set-up and the steady-state. The CHs are selected and the clusters (collection of nodes) are formed during the first stage. Afterward, a CH announces itself to the cluster members and assigns a time division multiple access (TDMA) table to each member in a cluster [65]. Figure 3 shows the set-up phase in LEACH.

To select a CH [66,67], nodes in each cluster create random numbers between 0 and 1 in each round. A reference number $T(n)$ is predefined to compare with the created numbers. In which the value of the reference number is greater than the value of the randomly created number, then the node that created the random number becomes the CH. Equation 1 is used to determine the value of the reference number $T(n)$ [65]:

$$T(n) = \begin{cases} p \times \left( r \mod \frac{1}{p} \right) & : \text{if } n \in G \\ 0 & : \text{otherwise} \end{cases}$$

(1)
where \( p \) is the percentage of CHs, \( r \) is the existing round, and \( G \) is a group of nodes that did not become CH in the prior \( 1/p \) rounds. In this manner, the energy is distributed among all nodes in the cluster. After CHs are chosen, a decision will be taken by each normal node to decide which CH is suitable to join with to form clusters \([68]\). CH sends an advertisement message to all other nodes. After that, each node selects the appropriate CH to connect with it depending on the strength of a signal, if the advertising CH is close to a node, then the signal has high strength and vice versa. When a node selects its potential CH, it transmits a message to that CH to be a member of a cluster. Then, the CH sends a timetable to other member nodes in its cluster.

The table is used to avoid data collision during transmission \([65–69]\). During the steady-state stage, each CH collects data, which are sent by sensor nodes in its relative cluster, and transmits them to the BS \([70]\). The LEACH protocol has many advantages \([71]\). It balances power consumption among all nodes in each cluster due to the rotation of the CH selection process. This procedure can increase the networks lifetime.

Furthermore, using the TDMA table allows nodes to be in sleep mode while waiting for their turn to send their data; this condition also increases the network lifetime. By contrast, several disadvantages exist. For example, LEACH does not take into account the residual energy of a node in the selection process of a CH. According to Equation (1), the possibility of a node becoming a CH is not fixed. After several rounds, all nodes have the same probability to become CHs. Thus, nodes with a high amount of energy and nodes with a low amount of energy have an equal chance to become a CH.

Another problem is the direct transmission from a CH to the BS using single-hop communication. This situation would be a serious issue if the CH and the BS were located in far positions because it uses more energy than a CH that is close to the BS, resulting in energy holes in which isolated nodes are unable to transmit their data. This situation considerably affects the performance of the network \([72,73]\). To overcome these disadvantages of LEACH, many improvements have been proposed by modifying LEACH to increase efficiency, as presented in Section 2.3.2.

2.3.2. Overview of Several Existing Surveys

Many surveys \([30–33]\) have been conducted to investigate the LEACH routing protocol and its variants for WSNs. In \([30]\), Sunil et al. conducted a review that discussed the successors of the LEACH routing protocol. The authors classified protocols by using the category of data transmission. Moreover, the paper presented more than 60 LEACH variants and discussed the advantages and drawbacks of each routing protocol. The authors compared the protocols based on nine different parameters. However, the limitation of \([30]\) is that the used classification was based only on single-hop and multi-hop communication without considering any other categories.

In \([31]\), Maurya and Kaur provided a short overview of the LEACH protocol and its improvements. The protocols were compared in terms of node mobility, location awareness of nodes, and method of data transmission between CHs and the BS. The limitation of \([31]\) is that the authors discussed routing protocols based on LEACH without using any classification. In \([32]\), Vishal et al. focused on the descendants of the LEACH protocol that overcome the non-uniform distribution of CHs in LEACH. This work was extended to investigate other hierarchical routing protocols and compare them with the LEACH protocol depending on the energy consumption. The limitation in \([32]\) is that the authors only discussed six variants of the LEACH protocol without any classification. They merely presented the variants as solutions of the non-uniform distribution of CH in LEACH.

In \([33]\), Singh conducted a survey of various LEACH-based protocols. The author selected several LEACH-based protocols and conducted a structural analysis of them. In addition, the survey in \([33]\) compared routing protocols on the basis of many comparison metrics, such as mobility, scalability, hop limit, energy efficiency, and use of localization information. Meanwhile, the survey \([33]\) discussed only sixteen protocols without any classification; it merely presented the protocols briefly. A comparison of existing surveys is shown in Table 3.
Table 3. Comparison of previous surveys on leach-based routing protocols.

<table>
<thead>
<tr>
<th>Previous Survey</th>
<th>Contributions</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>[30]</td>
<td>Sixty LEACH variants protocols are presented and compared using eight parameters.</td>
<td>Only communication–related protocols are presented based on single-hop and multi-hop communication.</td>
</tr>
<tr>
<td>[31]</td>
<td>Many LEACH–based protocols are discussed and compared using three parameters.</td>
<td>Protocols are discussed without any classification, and the authors used only three parameters for comparison.</td>
</tr>
<tr>
<td>[32]</td>
<td>Six protocols are discussed to overcome the non–uniform distribution of CH in LEACH.</td>
<td>Only six LEACH variants are presented without any classification.</td>
</tr>
<tr>
<td>[33]</td>
<td>Sixteen LEACH variants are discussed and compared using seven parameters.</td>
<td>Only 16 LEACH variants are discussed briefly without any classification.</td>
</tr>
</tbody>
</table>

In this survey, a clear and comprehensive classification of LEACH variants is proposed with a focus on the criteria that are used to select a CH and the criteria that are used to form clusters in LEACH–based protocols. This survey is completely different from the above mentioned surveys that focused either on a certain classification of LEACH–based protocols or on the general design matters and performance improvements of LEACH–based routing protocols.

2.3.3. Proposed Taxonomy of LEACH-based Routing Protocols in WSNs

In the current survey, the focus is on CH selection and cluster formation techniques as designated criteria to present the classification of LEACH–based protocols in WSNs. In Figure 4, LEACH–based routing protocols are categorized into two categories: CH selection-based and cluster formation-based protocols. Each main category is further classified into different classifications based on the main factor in the proposed approaches.

![Figure 4. Proposed taxonomy of LEACH–based routing protocols in WSNs.](image-url)

The next subsections describe each main category in detail. Then, existing familiar protocols that belong to each main category are presented in detail.
2.3.4. CH Selection-based Routing Protocols

In routing protocols, selection of the CH is a significant criterion to maximize network lifetime [74–76]. The main idea in all CH selection-based routing protocols is to choose the most efficient node to be the head of a cluster [77,78]. These protocols consider specific metrics to choose the suitable CH. Based on the selection mechanism of the CH, these protocols are classified into three classes, namely, residual energy, energy efficiency, and distance. The following paragraphs discuss these categories in detail as regards CH selection.

Residual Energy

The CH selection process in the clustering protocols can be conducted depending on the lasting energy of each node in a cluster. Residual energy exerts a major impact on the performance of a network and the networks lifetime. Specifically, if the node selected to be a CH has the highest level of energy among all nodes in the cluster, this will maximize the network lifetime, and the probability of a CH dying early will decrease. The subsequent paragraphs discuss the protocols that belong to this group in detail.

The centralized cognitive LEACH (CogLEACH–C) protocol overcomes a major problem in WSNs, namely, the residual energy in a CH [79]. A centralized spectrum-aware algorithm for the cognitive radio sensor network (CogLEACH–C) is used to address the level of energy to select a CH, in addition to the number of idle sensed channels. This protocol is centralized. Thus, the BS is responsible for determining if a node will become a CH or not. First, each node transmits a message to the BS with its ID, its position in the network, and its residual energy. Second, the BS computes the total remaining energy and how many idle channels in a network. Third, the decision for each node to be a CH or not is determined by the BS.

The major advantage of the CogLEACH–C protocol is that network lifetime is significantly increased because of the mechanism used to select a CH. Thus, the node with a high level of energy has a high probability to become a CH, thereby prolonging network lifetime. However, CogLEACH–C suffers from disadvantages. It uses location information metrics in messages sent to the BS to select the CH node; this process is costly because of the usage of GPS. Moreover, the nodes’ residual energy is the key metric in CogLEACH–C. Therefore, CH selection is directly affected by the total residual energy in the network. CogLEACH–C also relies on the BS in calculating the probability of nodes to become CHs. Therefore, message overhead in the network increases because of the information sent from each node to the BS.

Another protocol in this subcategory is vice CH (V–LEACH) [80]. Different from LEACH, this protocol uses a vice CH node via the residual energy of nodes, thereby increasing the efficiency of selecting the CH. Figure 5 shows the architecture of the V–LEACH protocol.

![Figure 5. V–LEACH architecture [80].](image-url)
In V–LEACH, a new mechanism is proposed to avoid the early death of CHs due to consuming more energy than member nodes do. Thus, nodes are reorganized and CHs are reselected based on the vice CH (VCH) concept [81–83]. When the CH has the lowest level of energy, it selects a VCH depending on the residual energy and then sends the VCHs ID to all other nodes in a cluster. Thus, normal nodes will begin to transmit data to the VCH instead of the previous CH, which will become a standard node.

In V–LEACH, network lifetime is significantly increased due to selecting a vice CH instead of a re-clustering operation, thereby saving energy in the entire network. Moreover, the efficiency of sending data to the BS is increased because the data are sent to the BS even the death of a CH by using the vice CH. Thus, sending packets to the BS is guaranteed in this protocol. However, the transmission cost is not considered in this protocol.

The improved LEACH (I–LEACH) protocol concentrates on the lasting energy of sensor nodes in a cluster [84]. In the LEACH protocol, a CH is selected depending on a predefined threshold without considering the remaining energy of a node. This approach has a seriously negative influence on the network lifetime when a little energy node becomes a CH. The threshold formula in LEACH is updated by introducing the remaining energy factor, as presented in Equations (2) and (3) [84]:

\[
T(i) = \begin{cases} 
1 - P_i \times \left( r \times \text{mod} \left( \frac{1}{N} \right) \right), & N \in G \\
0, & \text{otherwise}
\end{cases}
\] (2)

where:

\[
P_i = \frac{(E_i - E_r)^2}{E_R}
\] (3)

where \(E_i\) and \(E_r\) denote the lasting energy in each sensor and the average energy in the rest of sensors in round \(i\), respectively. \(E_R\) denotes total energy in the remaining sensors. To calculate \(E_r\), Equation (4) is used:

\[
E_r = E_R \times \left( 1 - \frac{r_i}{r_{\text{max}}} \right) / N
\] (4)

where \(r_i\) is the current round and \(r_{\text{max}}\) is the maximum rounds of network simulation.

According to Equation (3), the value of \(P_i\) will increase if much energy is present in a node. This increase will affect \(T(i)\) in Equation (2). Then, the probability of such nodes to become CHs is also increased.

If the chance of a node to become a CH increases and the lasting power in this node is high, network lifetime would likely increase. Similar to the abovementioned protocols, I–LEACH is better than LEACH in increasing network lifetime and decreasing power consumption. However, this protocol does not consider the transmission cost problem.

A new improvement method to increase network lifetime in a multi-hop network called LEACH one round (LEACH–1R) was proposed in [85]. This protocol changes a CH to another node when it runs out of energy instead of changing in each round as in LEACH. When the energy of the existing CH is below the threshold, it finds a node with more energy and informs it to become a new CH by sending an informing message. After receiving the informing message, the new CH sends the message that contains its ID, location, and energy level, to other nodes in the cluster.

This protocol has good performance in term of residual energy efficiency. However, the main problem is the changing of a CH after its energy is consumed; this may cause an energy hole after several rounds. Moreover, this protocol does not introduce a solution to the transmission cost problem. A summary of these protocols based on residual energy in selecting a CH is presented in Table 4. The strengths and weaknesses of the protocols are included.
Table 4. Summary of residual-based CH selection protocols.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>CogLEACH–C [79]</td>
<td>Balances energy load in a network via the use of idle channels with residual energy.</td>
<td>Inappropriate for wide networks.</td>
</tr>
<tr>
<td>V–LEACH [80]</td>
<td>Ensures data delivery because it uses two CHs.</td>
<td>Uses an extra CH which leads to more overhead, single hop communication.</td>
</tr>
<tr>
<td>I–LEACH [84]</td>
<td>Minimizes average energy consumption for each sensor node.</td>
<td>Computation overhead.</td>
</tr>
<tr>
<td>LEACH–1R [85]</td>
<td>Network lifetime and energy consumption.</td>
<td>Energy hole problem because of changing a CH after it runs out of energy.</td>
</tr>
</tbody>
</table>

Energy Efficiency

One of the basic issues in routing protocols is selecting a CH with the highest energy efficiency. This selection exerts a significant influence on the overall network lifetime. A CH selection-based routing protocol that concentrates on residual energy is not energy-efficient because it does not consider other factors, like the communication cost, energy consumption, and energy harvesting.

Moreover, many existing routing protocols select the CH depending on residual energy. Specifically, CogLEACH–C, V–LEACH, I–LEACH, and LEACH–1R use only a residual energy parameter. EEM–LEACH and EMOD–LEACH use a residual energy parameter with a transmission cost metric, EHA–LEACH uses a residual energy factor with energy harvesting capacity, and LEACH–MAC uses a residual energy criterion with the number of CH advertisements. The following paragraphs present routing protocols that emphasize the energy efficiency factor for selecting CHs.

To improve CH selection in LEACH and enhance the transmission of data, a new protocol called energy efficient multi-hop (EEM–LEACH) was proposed [86]. To select a CH, a threshold is calculated by specifying the residual energy. The new equation of the threshold value is:

\[
T(n) = \begin{cases} 
\frac{p}{1 - P \times (r \mod 1)} \times P(RE) : & \text{if } n \in G \\
0 : & \text{otherwise}
\end{cases}
\]

where \(P(RE)\) is the residual energy of the node and is calculated as follows:

\[
P(RE) = \begin{cases} 
\frac{E_{res} - E_{avg}}{E_{res}} : & \text{if } E_{res} > E_{avg} \\
1 - E_{avg} : & \text{otherwise}
\end{cases}
\]

In this protocol, the communication cost is computed by each CH when a CH advertisement is received from the BS. The communication cost of node \(u\) can be calculated as follows:

\[
c = \begin{cases} 
C(n) + d_{nm}^4 : & d_{nm} \geq d_0 \\
C(n) + d_{nm}^2 : & d_{nm} < d_0
\end{cases}
\]

where \(d_{nm}\) denotes the distance from node \(n\) to node \(m\) and \(C(n)\) helps in constructing a path from a CH to the BS.

This enhancement protocol provides an efficient technique for reducing energy consumption and increasing packet delivery. However, packet delay increases significantly because of the calculations in computing the communication cost for each node to select a CH and in computing the nodes remaining energy for specifying the nodes in the path. Moreover, the distance from the BS to a CH is not considered.

Enhanced modified LEACH (EMOD–LEACH), a new method that implements the idea of effective CH replacement and double transmitting power with an intracluster transmission manner, was proposed in [87]. To replace a CH, the algorithm checks the energy level in the current CH. If the
level is greater than the specific value, then the current CH will still the head in the following round. Otherwise, the set-up phase will be run again, and the current CH will be replaced to minimize power consumption.

The proponents of this protocol proposed two specified values (hard and soft) to minimize data consumption in redundant value sensing in addition to the residual energy. Data transmission process in this method depends on the value of the sensed event if this value is greater than the predefined hard value, and the predefined soft value is less than the difference between the current sensed and the previously sensed values; then the sensor node transmits the sensed data.

The EMOD–LEACH protocol is effective for applications that the time is important in them, but the amount of transmitted data to a CH is less than that in LEACH. Furthermore, the distance problem persists in this protocol. Different from EMOD–LEACH, a new protocol that selects a CH in WSNs has been designed based on the idea of using energy harvested sensor nodes; this protocol is called energy-harvested-aware LEACH (EHA–LEACH) [88].

In this protocol, a node with a high capacity for energy harvesting has a high probability to be a CH. The following formula calculates the total energy harvested in a time interval [0, \( T \)] [88].

\[
E_h(m, 0 \leq t \leq T) = \beta \ast \int_0^T P_h(m, t) dt - \int_0^T P_{\text{leak}}(m, t) dt
\]  

(8)

where \( P_h(m, t) \) indicates the harvesting rate of node \( m \) during time \( t \) and \( P_{\text{leak}}(m, t) \) is the leakage power of node \( m \) during time \( t \). \( T \) is a positive time scale. The proposed function in this protocol to measure the energy harvesting of node \( m \) can be presented as [88]:

\[
F(m) = e^{\frac{(E(m,0)+E_h(m,0 \leq t < T)-M)}{A}}
\]  

(9)

where \( e \) is the base of the natural logarithm, \( A \) is the variance value of the energy of each node in the network and \( M \) is the mean value. To select a CH, each node follows the same procedure that is described before when explaining the LEACH protocol to create a random value. Afterwards, a node matches this value with a reference number; if it is less than the reference number, then this node becomes a CH. The reference number can be computed as follows:

\[
T(m) = \frac{F(m)}{\sum_{v \in L(m)} F(v)} \times p \times |L(m)|
\]  

(10)

where \( p \) is the percentage of the CH nodes and \( L(m) \) is the set of the neighboring nodes of node \( m \).

This protocol outperforms LEACH protocol based on the energy consumption factor, but the main drawbacks of it are the cost and overhead. Moreover, this protocol does not present a solution to the distance problem. The LEACH medium-access control routing scheme (LEACH–MAC) was presented in [89]. This protocol aims to prevent the problem of random selection of CHs in LEACH because this problem leads to decrease the network lifetime and in-crease energy consumption.

This protocol resolves the previously mentioned drawback by limiting the CH advertisements in a network. When a CH selection begins, a new variable called \( CH_{\text{heard}} \) [89] starts with 0 and increases when an advertisement message is received from a CH. The sending time of a CH, which is the time required for sending and receiving operations performed by it, is in the interval [0 to total advertisement time]. Each node selects a random number from this interval in the threshold function. If the time selected by a node is \( R_t \), then the advertisement sending time of the cluster head is calculated in Equation (11):

\[
l_{\text{adv–CH}} = \frac{R_t}{\text{Current Energy}}
\]  

(11)
If a number of clusters is greater than the value of $CH_{heard}$, then the node announces itself as a CH. Next, it will circulate an advertisement note to other nodes. The total energy consumed by a CH in receiving, collecting, and transmitting data (ECHMAC) is calculated as follows:

$$E_{CHMAC} = lE_{elec} \frac{N}{A} + lE_{DA} \frac{N}{A} + l_{\text{emp}}d_{BS}^4$$  (12)

where $A$ denotes CH advertisements, $l$ is the length of data, $E_{elec}$ is the consumed energy, and $d$ represents the distance to the BS. The energy consumed by other nodes that are not CHs is found using the following equation:

$$E_{Non-CHMAC} = lE_{elec} + l_{\text{fs}} \frac{M^2}{2\pi A}$$  (13)

Similar to LEACH, this protocol increases network lifetime, but it possesses high complexity due to the calculations and overhead. A summary of these protocols based on energy efficiency in selecting a CH is presented in Table 5. The strengths and weaknesses of each protocol are also provided.

### Table 5. Summary of energy efficiency-based CH selection protocols.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>EEM–LEACH [86]</td>
<td>Energy consumption and packet delivery</td>
<td>Packet delay increases because of computing the residual energy for relay nodes</td>
</tr>
<tr>
<td>EMOD–LEACH [87]</td>
<td>Efficient CH replacement scheme enhances network lifetime</td>
<td>Unsuitable for periodically sensing data</td>
</tr>
<tr>
<td>EHA–LEACH [88]</td>
<td>Using energy harvesting nodes increases network lifetime</td>
<td>High complexity</td>
</tr>
<tr>
<td>LEACH–MAC [89]</td>
<td>Energy efficiency due to consider the number of CH advertisements.</td>
<td>Complexity and extra overhead.</td>
</tr>
</tbody>
</table>

**Distance**

The distance from a CH to the BS is a critical metric in designing routing protocols. A problem occurs when a CH is far from the BS. Given that WSNs suffer from dynamic and reselecting CHs in each round, distance is important in selecting a CH due to the main function of a CH, which is sending aggregated data to the BS. Thus, the distance will affect the process of data transmission.

The following paragraphs discuss CH selection-based protocols that focus on the distance factor. Improved-LEACH is another enhancement of LEACH, and it is one of the protocols based on distance when selecting a CH [90]. Two factors are considered in this protocol when selecting a CH: the level of energy in nodes and the distance from each node to the BS. Accordingly, Equations (14) and (15) are used to find a lasting energy for each node in a cluster, and the distance to the BS, respectively:

$$SCRFN1 = \frac{E_{current}}{E_{max}}$$  (14)

where $E_{current}$ denotes the energy level in a node and $E_{max}$ represents the maximum energy, and:

$$SCRFN2 = \frac{d_{bs}}{d_{far}}$$  (15)

where $d_{bs}$ denotes the distance from a node to the BS and $d_{far}$ denotes the distance from the farthest node in a cluster to the BS.
Both functions in Equations (14) and (15) are then combined and multiplied by the probability function to prolong the networks lifetime. Equation (16) calculates the new value of selecting a CH:

\[
SCFN = \left[ w_1 \left( \frac{E_{current}}{E_{max}} \right) + w_2 \left( \frac{d_{bs}}{d_{far}} \right) \times \left( \frac{p}{1 - \frac{p}{r \mod \frac{1}{p}}} \right), \text{if } n \in G \right], \text{otherwise} \]

(16)

where \( w_1 \) and \( w_2 \) are weights that are given according to the preference for energy and distance, where \( w_1 + w_2 = 1 \).

This protocol enhances the network efficiency in comparison with the LEACH protocol. In addition, the number of alive nodes is increased when Improved–LEACH is used. However, the main problem with this protocol is the calculation overhead.

A new routing protocol based on annealing and genetic algorithms called LEACH simulated annealing and genetic algorithm (LEACH–SAGA) was proposed in [91] to enhance energy distribution among nodes in the network. In LEACH–SAGA, clusters are formed by using genetic and annealing algorithms. Afterward, the BS finds the center point in each cluster, and then, it considers the energy level and the distance from each node to the center of each cluster.

Average energy in each cluster is calculated and the remaining energy in each member node will be compared with the average. All nodes that have energy more than the average will be nominated as CHs of that cluster. Then, the final decision of selecting the potential CH will be taken by the BS depending on the distance from the center. Therefore, the distance plays a major role in this protocol.

LEACH–SAGA protocol is better than LEACH protocol based on network lifetime. Moreover, this protocol solves the randomization problem in the LEACH protocol, which is discussed before, by using the distance factor. The main disadvantage of this protocol is the complexity due to the use of a genetic algorithm. A summary of these protocols based on distance in selecting a CH is presented in Table 6. The strengths and weaknesses of each protocol are also provided.

Table 6. Summary of distance-based CH selection protocols.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved–LEACH [90]</td>
<td>Good selection of CHs and improvement of network performance</td>
<td>Computation complexity</td>
</tr>
<tr>
<td>LEACH–SAGA [91]</td>
<td>Efficient distribution of CHs by reducing energy consumption</td>
<td>Scalability and delay</td>
</tr>
</tbody>
</table>

2.3.5. Cluster Formation–Based Routing Protocols

Cluster formation–based routing protocols do not fully concentrate on selecting CHs. Routing protocols in this category use some information to form clusters in WSNs. Based on the formation mechanism of clusters, these protocols can be divided into five categories: residual energy, energy efficiency, network coverage, routing path, and distance. The following subsections present these subcategories and discuss the corresponding protocols in detail.

Residual Energy

In this subcategory, the residual energy of nodes is considered to increase the network lifetime. Following paragraphs discuss the protocol based on residual energy to form clusters in detail. The authors in [92] presented a protocol called energy efficient LEACH (EE–LEACH) to provide the best cluster formation and data aggregation techniques.

The clusters in this protocol are formed by using the information of neighbor nodes and their level of residual energy. In this protocol, the information of neighbor nodes is taken into account when forming clusters. The list of all neighbor nodes is sorted in descending order depending on the hop distance. Then, all members of that list are analyzed, and any two-hop neighbor is a candidate
for the cluster. Afterward, the neighbor nodes will be sorted according to their residual energy by executing a sorting algorithm. This protocol has several advantages. For example, it selects nodes with the highest remaining energy to send data to the BS, thus increasing the network lifetime. Moreover, this protocol provides a good ratio of packet delivery. However, complexity increases due to the use of several techniques.

Energy Efficiency

This factor is important because of its influence on the design of routing protocols. The main goals of protocols belonging to this subcategory are to prolong network lifetime and minimize energy consumption by using factors that help in achieving these goals. The following paragraphs discuss these protocols in detail. The intra balanced LEACH protocol (IB–LEACH) [93] was presented to minimize the energy consumed between sensor nodes and their CH inside a specific cluster and to reduce the load on a CH by distributing its tasks among all nodes in the cluster.

IB–LEACH has several rounds as mentioned before in the LEACH protocol, but in IB–LEACH, each round comprises three stages. Set-up and steady–state stages are similar to the stages in LEACH. In the new stage, which is called the pre-steady, nodes in each cluster are tagged as CH, sensor nodes, and aggregator. Sensor nodes sense events and then transmit data to the aggregator. The main jobs of the aggregator node are the aggregation of data and the transmission of them. The task of a CH is to manage the activities, such as creating the TDMA schedule and distributing it among nodes. In this manner, the tasks are distributed among all nodes in the cluster, which results in minimizing the load on the CH.

The main advantages of the IB-LEACH protocol are as follows: (1) network lifetime significantly improves (2) and energy consumption declines because of the identical distribution of energy in a cluster. However, the drawback of this protocol is the extra overhead in messages that are circulated to select CHs and aggregators.

To equally split the energy among nodes in wide area networks, and to increase the performance of data transmission, Partition–LEACH was presented in [94]. In this protocol, the formation of the cluster is done in circle shapes with equal radius for each, this circle-shape of clusters increase the transmission efficacy. The first cluster is formed and a sensor node that has the maximum energy level becomes the center node of it. Other clusters will be formed around the first one, and the node that is positioned in the center of each cluster becomes the center node of it. The overall structure of a network in this protocol consists of partition clusters, the communication region, and the four partition clusters structure. Each partition cluster has a central nod at the center of it, and four gates nodes to transmit data and monitor the location of mobile sink.

If the mobile sink is detected by one of the gate nodes, it sends a message to the others to convert their status to sleep. After that, they will be in a ready state and waiting for a message to wake up in case the mobile sink reaches close to the partition cluster. When they receive a message, they will start sending data to the mobile sink via one of the gate nodes. After finishing the transmission process, and when the mobile sink leaves the partition cluster, nodes will go back to the deep sleep status.

Comparable to IB–LEACH, Partition–LEACH increases network lifetime because of the equal split of energy. Furthermore, this protocol is suitable and efficient for large area networks due to the good performance in terms of data transmission.

The main drawback of Partition–LEACH is the extra overhead in messages due to the existence of the mobile sink. A new method to reduce fault tolerance in WSNs, called fault tolerance LEACH (FT–LEACH) [95], has also been proposed. Fault tolerance is a problem in LEACH protocols because it affects performance. Different from IB–LEACH and Partition–LEACH, this method concentrates on fault tolerance to increase efficiency.

Depending on the energy values, the fault is detected. FT–LEACH considers the cluster formation in two ways; global and local re–clustering. The local re–clustering will be performed to repair the cluster, and global re-clustering is performed to repair the network. If the fault is in a CH, local
re-clustering will be performed after deleting the CH. This protocol is better than LEACH based on energy consumption and fault tolerance.

A summary of these protocols based on energy efficiency in cluster formation is presented in Table 7. The strengths and weaknesses of each protocol are included.

**Table 7. Summary of energy efficiency-based cluster formation protocols.**

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>IB-LEACH [93]</td>
<td>Uniform energy distribution in each cluster.</td>
<td>Control message overhead for selecting CHs and aggregators.</td>
</tr>
<tr>
<td>Partition-LEACH [94]</td>
<td>Balanced distribution of energy among nodes and increase data transmission efficiency.</td>
<td>Extra overhead because of the mobility of the sink.</td>
</tr>
<tr>
<td>FT-LEACH [95]</td>
<td>Fault tolerance</td>
<td>Re-clustering and duplicate data management are not discussed.</td>
</tr>
</tbody>
</table>

**Network Coverage**

Network coverage is important when designing routing protocols. It is necessary for each routing protocol to cover the entire network and efficiently distribute the nodes in the target space to achieve this type of coverage. The protocols that depend on network coverage are discussed in detail in the following paragraphs. In [96], LEACH and PEGASIS protocols were combined to develop a new reliable protocol for routing data in WSNs called P-LEACH. The idea of data transmission in clusters was utilized along with chain formations that are used in LEACH and PEGASIS to increase the network coverage.

In P–LEACH [96], the node with the maximum level of energy is selected by other nodes to be a CH. After determining the CHs in each cluster, a chain is created with the CHs. The nearest CH to the BS becomes the leader of the chain. When a CH collects data that are sent by other nodes in a cluster, it transmits these data to the next CH in the chain until it reaches the BS. In P-LEACH, if a CH has minimum energy and cannot send data, then the node with the second highest value of energy will be a new CH to ensure efficient operation in the network and keep the coverage of the network stable. The integration of the best features in LEACH and PEGASIS into the P-LEACH protocol leads to good performance and improved network coverage in WSNs.

Another protocol in this subcategory is the Cell LEACH protocol (C–LEACH) [97], which was proposed to increase network coverage in WSNs. Different from P–LEACH, C–LEACH divides the entire network into sets of cells with many sensor nodes in each cell. Every adjacent seven cells are grouped together and form one cluster. In each cell, there is a private node named the cell head. The function of this node is to allocate a time slot using TDMA to cell members, aggregate data from other nodes, and send these data to the respective CH. The same thing is performed to transmit data to the BS using the optimal route.

This protocol has many advantages. It has good network coverage due to the technique of dividing networks into cells. Moreover, it increases the energy efficiency. The disadvantages of this protocol are the increased of control packet overhead and the complexity due to the different head nodes in this protocol, which are cell heads and cluster heads. Another protocol that provides high network coverage is Orphan LEACH (O–LEACH) [98].

Figure 6 shows the technique used in this protocol to partition the network into cells and to form a cluster with seven nearby cells.
Sensor nodes are connected to CHs in each cluster, but several nodes are still not connected. These nodes are considered orphan nodes. These nodes transmit data to gateway nodes. Afterward, the gateway nodes play the role of a CH by gathering and transmitting data to the BS. Another way that has been proposed in this protocol is to form a special cluster of orphan nodes and to select a private CH, which belongs to that special cluster.

The task of this CH is the same task as the traditional CH but in this case, the communication is done between the special CH and the gateway. This protocol provides great network connectivity and good coverage, but the problem is the delay in data delivery. The Energy Zone LEACH (EZ–LEACH) protocol [99] is an enhancement of the LEACH protocol.

This protocol differs from previous ones in which the location of nodes is utilized to select a CH after dividing the sensing area into several zones. The advantages of dividing a network into zones are to ensure the equivalence of distances from a CH to other nodes that belong to each zone and to select a fixed number of CHs according to the number of zones in each round.

To select a CH, the distance from each node to the center of the zone is the main used parameter. The BS selects nodes that have a minimum distance to the center in each zone. Then, all other nodes join with a suitable CH. Each node sends the information of its distance from the zone center with the energy level, and a new CH is selected according to Equation (17):

$$Z = \{z \mid \forall \text{sensor} z \in \text{cluster}, E_{\text{sensor}z} \geq \text{Thre Value}\}$$

$$\forall z \in Z : \text{if } d_z < d_{z+1}$$

Index $\leftarrow z$

New Cluster $\rightarrow$ Head $\leftarrow$ Sensor Index

where sensor $z$ is any sensor in a cluster and $E_{\text{sensor}z}$ is the remaining energy in it. $z, z + 1,$ and index are indices.

The EZ–LEACH protocol prolongs networks lifetime and network coverage due to the efficient way of forming clusters, that depends on the number of zones. Summary of these protocols based on network coverage in cluster formation is presented in Table 8. The strengths and weaknesses of each protocol are also provided.
Table 8. Summary of network coverage-based cluster formation protocols.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>O–LEACH [98]</td>
<td>Good connectivity and coverage.</td>
<td>Delay of data delivery and control overhead</td>
</tr>
<tr>
<td>EZ–LEACH [99]</td>
<td>Efficient CH selection based on different zones.</td>
<td>Extra overhead.</td>
</tr>
</tbody>
</table>

Routing Path

The process of sending the data from a source to a destination using an appropriate path is called routing. In WSN, many protocols have been proposed to select an efficient routing path in order to increase the network lifetime. The main issue in routing protocols is the selection of the efficient routing path. Sending data from a CH to the BS using an efficient way is important in prolonging the network lifetime and reducing the power consumption during transmission. The following paragraphs discuss in detail the cluster formation-based routing protocol belonging to this subcategory.

A protocol called Cross-Layer LEACH (CL–LEACH) [100] has been proposed by utilizing the cross-layer technique to maximize the lifetime of the WSN. This protocol comprises four stages. In the first stage, which called cluster formation stage, a CH is selected depending on the distance to the BS and the energy level. In the second stage, a route will be created by using the routing mechanism. To find the route, distance is computed using the following equation:

\[ D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \]  

(21)

where \( D \) is the distance, \((x_1, y_1)\) are the coordinates of the source node, and \((x_2, y_2)\) are the coordinates of the destination node.

During the third stage, the residual energy of each node is compared with a threshold; if it is greater, then the node is considered to be a relay node in the path. If a route between the source and destination is damaged, then the route maintenance technique finds the broken node in a route and substitutes it with another one. This protocol is better than LEACH based on energy consumption and live nodes, but the overhead and complexity are its main drawbacks.

Distance

Distance is important when forming clusters in WSNs to maximize the network efficiency. When the distance from the CH in each cluster to the BS is short, then energy consumption decreases. Familiar protocols that depend on the distance factor when forming clusters are discussed in detail in the next paragraphs.

In this subcategory, the first protocol is the Dual-hop Layered LEACH (DL–LEACH) [101] protocol that uses the multi-hop routing technique in two layers according to the distance to the BS. Lower layer contains nodes that are near to the BS. When these nodes wish to send data, they compare the distance to the CH and to the BS. If the distance from a node to the BS is shorter than the distance from a node to the CH, then it will send data to the BS directly; otherwise, it will transmit data through the CH. Energy consumption is improved in this protocol, but in large networks, the lifespan of nodes is short. Another protocol that handles the distance factor in WSNs when forming clusters is Intermediate Cluster Head LEACH (ICH–LEACH) [102], in which the problem of LEACH related to the distance from a CH to the BS is considered.

As mentioned before, in the LEACH, data are sent directly from a CH to the BS regardless of the distance between them; hence, the necessary energy to send data will be very high if a CH and the BS
are far from each other, thus affecting the lifetime of the network noticeably. To overcome this problem, a new method has been proposed to maximize network lifetime and minimize power consumption.

The idea is to use an intermediate CH if the current CH locates far from the BS. The selection of an intermediate CH depends on the BS, CH, and intermediate CH positions. In the initialization stage, the BS transmits its coordinates to all nodes in the network, and these coordinates are used by CHs to determine if they can transmit data to the BS directly. In the cluster setup stage, each CH sends its distance from it to the BS with its position to other CHs. Such information is used to form a routing table in each CH to be used in the next stage. During the steady stage, a CH checks its routing table to decide whether it can send the collected data to the BS without any relay CH (if it is the nearest node) or through intermediate CH. The source CH selects an intermediate CH in a case wherein the distance from the intermediate one to the BS is less than the distance from the source CH to the BS. The ICH–LEACH protocol is good at extending lifetime of a network compared with the LEACH protocol, and the receiving packets by the BS increase. Moreover, the dead nodes and the total consumed energy decrease.

A summary of these protocols based on distance in cluster formation is presented in Table 9. The strengths and weaknesses of each protocol are also provided.

### Table 9. Summary of distance-based cluster formation protocols.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICH–LEACH [102]</td>
<td>Increases the lifetime of a network by using intermediate CH.</td>
<td>Extra overhead.</td>
</tr>
</tbody>
</table>

3. Comparative Analysis

Clustering is one of the most stimulating fields in WSNs due to the exceptional features and nature of WSNs. Routing protocols based on the LEACH protocol can be categorized into two main categories: CH-based and cluster formation–based protocols. The protocols belonging to the CH–based category are more effective than LEACH because selecting the CH is performed in an efficient manner and by using efficient techniques.

The LEACH protocol has several limitations in terms of CH selection and cluster formation. This study emphasis on the new taxonomy in classifying LEACH variants, which is based on CH selection and cluster formation methods. These variants were developed to maximize the network lifetime and reduce the consumption of energy.

The protocols discussed in this study exhibit better performance than LEACH when considering the selection of CH, the formation of clusters, and the consumption of energy. The discussed routing protocols are summarized in Table 10. In this table, the comparison among these protocols is shortened based on the new proposed categories and subcategories.

The meaning of the suggested columns in the table is clarified as follows:

- **Category, Subcategories, and Protocol**: This column lists the two main categories used in this survey and the subcategories belonging to each one. The set of protocols assigned to each subcategory is also listed with the corresponding reference number for each protocol.
- **Research Domain**: This column presents the domain of each protocol.
- **Modeling Parameters**: This column lists the main parameters used in each protocol, and each protocol selects CHs or forms clusters based on different parameters.
- **Contributions**: This column summarizes the main contribution of each protocol.

Table 11 provides a summary of the protocols that are discussed above. This summarization is based on different parameters that affect each protocol to help researchers and developers in future studies. These parameters are residual energy, location, complexity, multi-hop, reliability, and load balancing.
Table 10. Features of routing protocols in WSNs based on the LEACH protocol.

<table>
<thead>
<tr>
<th>Category and Subcategories</th>
<th>Protocol</th>
<th>Research Domain</th>
<th>Modeling Parameters</th>
<th>Contributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster Head-Based</td>
<td>CogLEACH-C [79]</td>
<td>Cognitive radio</td>
<td>Idle channels and residual energy</td>
<td>Use residual energy in addition to idle channels</td>
</tr>
<tr>
<td>Routing Protocols</td>
<td>V-LEACH [80]</td>
<td>Multiple cluster heads in Clusters</td>
<td>Residual energy</td>
<td>Select a vice cluster head in case of main cluster head failure</td>
</tr>
<tr>
<td></td>
<td>I-LEACH [84]</td>
<td>Energy efficient</td>
<td>Residual Energy</td>
<td>Update the threshold value by adding the remaining energy as a factor</td>
</tr>
<tr>
<td></td>
<td>LEACH-IR [85]</td>
<td>Cluster head changing</td>
<td>Residual energy</td>
<td>Change the cluster head to another node when it runs out of energy</td>
</tr>
<tr>
<td>Energy Efficiency</td>
<td>EEM-LEACH [86]</td>
<td>Energy efficient</td>
<td>Residual energy and communication cost</td>
<td>Select the node with maximum residual energy as the cluster head</td>
</tr>
<tr>
<td></td>
<td>EMOD-LEACH [87]</td>
<td>Cluster head changing</td>
<td>Energy level and communication cost</td>
<td>Efficient cluster head replacement and dual transmitting power</td>
</tr>
<tr>
<td></td>
<td>EHA-LEACH [88]</td>
<td>Energy harvesting</td>
<td>Energy harvesting capacity</td>
<td>Maximizing the minimum energy conversation</td>
</tr>
<tr>
<td>Distance</td>
<td>LEACH-MAC [89]</td>
<td>Medium access control</td>
<td>Advertisement time and distance</td>
<td>Decreases the randomness problem for clustering by limiting the number of advertisements generated by cluster heads</td>
</tr>
<tr>
<td></td>
<td>Improved-LEACH [90]</td>
<td>Energy efficient</td>
<td>Distance between node and BS</td>
<td>Two factors are considered in selecting a cluster head which increase the number of alive nodes</td>
</tr>
<tr>
<td></td>
<td>LEACH-SAGA [91]</td>
<td>Optimization</td>
<td>Residual energy and distance from centroid of a cluster</td>
<td>Simulated annealing and genetic algorithms are used to find the optimal number of cluster heads</td>
</tr>
<tr>
<td>Cluster Formation-Based</td>
<td>EE-LEACH [92]</td>
<td>Energy efficient</td>
<td>Residual energy</td>
<td>Gaussian distribution for node deployment, relay node selection</td>
</tr>
<tr>
<td>Routing Protocols</td>
<td>IB-LEACH [93]</td>
<td>Data aggregation</td>
<td>Consumed energy</td>
<td>Minimizing the load of a cluster head by even distribution of tasks among nodes in clusters</td>
</tr>
<tr>
<td></td>
<td>Partition-LEACH [94]</td>
<td>Network hierarchy</td>
<td>Distance from cluster heads to the BS</td>
<td>Using two-level hierarchies of clusters, even distribution of energy among nodes</td>
</tr>
<tr>
<td></td>
<td>FT-LEACH [95]</td>
<td>Network reliability</td>
<td>Residual energy</td>
<td>Considering global and local re-clustering when detecting fault in nodes</td>
</tr>
<tr>
<td>Energy Efficiency</td>
<td>P-LEACH [96]</td>
<td>Data transmission</td>
<td>Residual energy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C-LEACH [97]</td>
<td>Network division</td>
<td>Residual energy and shortest path</td>
<td></td>
</tr>
<tr>
<td></td>
<td>O-LEACH [98]</td>
<td>Network division</td>
<td>Residual energy and distance from a node to the gateway node</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EZ-LEACH [99]</td>
<td>Network division</td>
<td>Nodes’ locations</td>
<td></td>
</tr>
<tr>
<td>Network Coverage</td>
<td>CL-LEACH [100]</td>
<td>Cross layer</td>
<td>Residual energy and distance from a node to the BS</td>
<td>Routing layer and physical layer are merged with MAC</td>
</tr>
<tr>
<td>Routing Path</td>
<td>DL-LEACH [101]</td>
<td>Layered approach</td>
<td>Distance to the BS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ICH-LEACH [102]</td>
<td>Multiple cluster heads in clusters</td>
<td>Distance to the BS</td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 11. Comparison of routing protocols in WSNs based on different parameters.

<table>
<thead>
<tr>
<th>Category and Subcategories</th>
<th>Protocol</th>
<th>Residual Energy</th>
<th>Location</th>
<th>Complexity</th>
<th>Multi-Hop</th>
<th>Reliability</th>
<th>Load Balancing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster Head-Based Routing Protocols</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual Energy</td>
<td>CogLEACH-C [79]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>V-LEACH [80]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>I-LEACH [84]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>LEACH-IR [85]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Energy Efficiency</td>
<td>EEM-LEACH [86]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>EMOD-LEACH [87]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>EHA-LEACH [88]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Distance</td>
<td>LEACH-MAC [89]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Improved-LEACH [90]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>LEACH-SAGA [91]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cluster Formation-Based Routing Protocols</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual Energy</td>
<td>EE-LEACH [92]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Energy Efficiency</td>
<td>IB-LEACH [93]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Partition-LEACH [94]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>FT-LEACH [95]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Network Coverage</td>
<td>P-LEACH [96]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>C-LEACH [97]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>O-LEACH [98]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>EZ-LEACH [99]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Routing Path</td>
<td>CL-LEACH [100]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Distance</td>
<td>DL-LEACH [101]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>ICH-LEACH [102]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Selecting the best node to become a CH is a problem in routing protocols and basically affects the overall performance of the network. The selection of the best node to become a CH affects energy consumption and network lifetime due to the tasks of the CH in aggregating data and sensing them to the BS.

Another critical issue in designing routing protocols is cluster formation, which also influences network lifetime. Most LEACH–based protocols were proposed to deal with these problems. As shown in Table 10, the majority of existing CH selection-based protocols select the CH node based on residual energy. Specifically, CogLEACH–C, V–LEACH, I–LEACH, and LEACH–1R employ only the residual energy parameter.

EEM–LEACH and EMOD–LEACH use the residual energy parameter and the communication cost, and EHA–LEACH utilizes the energy harvesting capacity parameter. I–LEACH and LEACH–SAGA consider the distance to the BS and residual energy, and LEACH–MAC uses advertisement time and distance. Meanwhile, cluster formation-based protocols use some information for forming clusters and re-clustering.

For instance, EE–LEACH uses residual energy for distributing nodes. C–LEACH, O–LEACH, and EZ–LEACH employ the locations of nodes and residual energy to divide the network into regions. DL–LEACH and ICH–LEACH utilize distance to the BS, and CL–LEACH utilizes distance to the BS and residual energy. FT–LEACH utilizes fault tolerance in considering local and global re-clustering.

Selecting the best CH node and cluster formation exert a significant impact on the performance of the overall network in terms of network lifetime, reliability, energy consumption, and other criteria. In LEACH, the randomness of the selection of CH and the direct communication between the CH and BS are the main factors that affect network lifetime.

However, existing CH–based and cluster formation-based protocols in WSNs suffer from many problems based on several metrics. Therefore, designing a CH selection protocol that selects an efficient and reliable CH is important. Designing a cluster formation protocol that efficiently forms clusters to minimize energy consumption, minimize complexity, and ensure packet delivery is also crucial.

A comparison of this survey with existing surveys [30–33] is presented in Table 12.

<table>
<thead>
<tr>
<th>Criteria-based taxonomy</th>
<th>[30]</th>
<th>[31]</th>
<th>[32]</th>
<th>[33]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detailed presentation of different protocols</td>
<td>√</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Detailed classification</td>
<td>√</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Scope of the survey</td>
<td>CH selection protocols and cluster formation protocols</td>
<td>Single-hop and multi-hop protocols</td>
<td>General review of LEACH descendant protocols</td>
<td>General review of LEACH variants and other hierarchical protocols</td>
</tr>
<tr>
<td>Qualitative comparison among protocols</td>
<td>√</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Comparison with previous surveys</td>
<td>√</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

### 4. Future Issues

CH selection and cluster formation in WSNs are major challenges due to the important role of CHs in gathering and sending data to the BS. All routing protocols based on LEACH concentrate on maximizing network lifetime and minimizing energy consumption. To accomplish these goals, different techniques have been proposed to enhance the LEACH protocol. The comparative analysis and discussions indicate that several aspects of LEACH must be considered by researchers when...
designing enhanced CH selection or cluster formation methods. The future issues in designing these methods are as follows:

- **Complexity**: Complexity is important in designing protocols, and it directly affects the overall performance of protocols, especially in WSNs, due to the nature of sensor nodes (small memory and small computation capability). Therefore, designing a CH selection protocol with low complexity is important.

- **Number of Rounds**: LEACH involves several rounds, and each round consists of processes (e.g., forming clusters, selecting CHs, and data transmission). The number of rounds is a significant factor in routing protocols because it affects the performance of the network. Hence, designing CH selection routing protocols that consider the number of rounds will improve network lifetime.

- **Scalability**: LEACH suffers from communication problems, especially in large-scale networks, due to the direct communication from a CH to the BS. This problem directly affects network lifetime because CHs die early in large-scale networks. To deal with this problem, multi-hop protocols with the capability to find the optimal and shortest path are recommended.

- **Security**: Security problems are major challenges in LEACH and its variants. The CH is the main target of hackers due to its importance in gathering and sending data. Thus, the CH should apply encryption and authentication algorithms to avoid attacks. The problem here is designing a suitable encryption algorithm for sensor nodes that have a small memory and computation capability to minimize message overheads and reduce energy consumption according to the calculations in encryption algorithms.

- **Fault tolerance**: In WSNs, sensor nodes are deployed in a large area and not accessible by humans. Fault tolerance exerts a significant impact on the lifetime of the network. Therefore, efficient algorithms that deal with faults must be designed, especially if the fault is in cluster head. How the algorithm selects a new CH in the case of failure and how to perform re-clustering while minimizing the message overheads must also be investigated.

5. Conclusions

WSNs have elicited much research interest in the last decade due to their numerous applications. The structure of routing protocols in WSNs is important when designing them and these protocols can be categorized into many categories based on them, with the important category of routing protocols being clustering or hierarchy routing.

Clustering routing presents many advantages, such as network scalability, energy distribution among nodes, and reduced load on the nodes in a network. Given these advantages, many protocols that use clustering routing have been proposed.

The LEACH protocol is considered as a common hierarchy protocol; it reduces power consumption by distributing energy among nodes in a network, but it has many disadvantages in terms of CH selection, cluster formation, and energy efficiency. Several advanced protocols have been conducted to improve LEACH.

In this survey, a new taxonomy was proposed to categorize many LEACH variants. The new taxonomy is based on the selection methods utilized to select a CH and on the methods utilized to form a cluster. These two categories were examined in this survey for classifying LEACH-based routing protocols because of their importance and impact on the performance of networks. Moreover, these two categories are considered main operations in clustering protocols; therefore, discussing and presenting routing protocols that depend on these methods are important.

In addition to using these two categories as the main classifying factors, this survey considered other important subcategories for further classification to make the classification clear and comprehensive. These new subcategories are residual energy, energy efficiency, distance, network coverage, and routing path. Due to their strong influence on the network lifetime, these parameters are considered in classifying the CH selection and cluster formation methods. This survey also discussed
21 routing protocols based on LEACH, and these protocols were classified under the abovementioned categories and subcategories. The classification of these protocols was performed in a tabular format for easy referencing. Furthermore, a qualitative comparison of this survey with other existing surveys was conducted, with a focus on contributions and limitations.

Moreover, future issues in proposing and developing new LEACH-based routing protocols were highlighted in this survey to help researchers by providing them useful information.

Author Contributions: The contributions of authors are as follows: Conceptualization: M.A.-S. and M.A.; formal analysis: M.A.-S.; investigation: M.A. and T.C.W.; project administration: M.A.-S.; resources: M.A.-S.; supervision: M.A. and T.C.W.; validation, M.A., T.C.W., and A.K.; writing—original draft: M.A.-S.; writing—review and editing: M.A. and A.K.

Funding: This grant is supported in part by Universiti Sains Malaysia through Grant No. 304/PNAV/650817/C127.

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

References


70. Wu, W.; Xiong, N.; Wu, C. Improved clustering algorithm based on energy consumption in wireless sensor networks. IET Netw. 2017, 6, 47–53. [CrossRef]

© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).