We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

6,600 Open access books available
177,000 International authors and editors
195M Downloads

154 Countries delivered to
TOP 1% Our authors are among the most cited scientists
12.2% Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Chapter

Data Collection Protocols in Wireless Sensor Networks

Koppala Guravaiah, Arumugam Kavitha and Rengaraj Leela Velusamy

Abstract

In recent years, wireless sensor networks have became the effective solutions for a wide range of IoT applications. The major task of this network is data collection, which is the process of sensing the environment, collecting relevant data, and sending them to the server or BS. In this chapter, classification of data collection protocols are presented with the help of different parameters such as network lifetime, energy, fault tolerance, and latency. To achieve these parameters, different techniques such as multi-hop, clustering, duty cycling, network coding, aggregation, sink mobility, directional antennas, and cross-layer solutions have been analyzed. The drawbacks of these techniques are discussed. Finally, the future work for routing protocols in wireless sensor networks is discussed.

Keywords: wireless sensor networks, routing protocols, data collection, network lifetime, energy efficiency, fault tolerance, low latency

1. Introduction

Wireless sensor networks (WSNs) [1] are distributed among environment with lightweight and small sensor nodes. These sensor nodes are used to measure the parameters of environment. Some of such parameters are vibration, pressure, sound, movements, temperature, humidity, etc. The sensors are well coordinated and connected to the base station (BS) or sink using wireless communication for forwarding sensed information. Due to this, many IoT-based applications such as home applications [2], vehicular monitoring [3], medical applications, structural monitoring, habitat monitoring, intrusion detection, tracking for military purpose, etc., are using WSNs for data collection [1, 4, 5].

Ad hoc and cellular network routing protocols are not suitable for sensor networks due to the sensor node design challenges such as node deployment, node mobility, and limited resource constraints (battery, communication, and processing capabilities) [6]. In WSNs, large number of sensor nodes are deployed for specific application due to this global addressing which is too difficult to maintain. Due to this large number, nodes located in the same area may generate redundant data and transmit to BS. This leads to bandwidth wastage and network traffic which in turn effects the more energy consumption. Another main resource constraint of a sensor node is limited battery power due to battery replacement or recharge not being possible in most of the WSN applications. WSN has a wireless communication medium, which
leads to an increased probability of collisions in the data communication process and which impacts on the network performance. While designing a new data collection routing protocol and achieving its requirements such as coverage area, data accuracy, and low latency, we need to consider the above stated issues [7].

In WSN, collection of sensed data can be done in a regular or non-regular mode. Data have to be collected continuously from sensor nodes in regular mode. Whereas, in the non-regular mode, the data have to be collected at some periodic intervals from sensor nodes.

Table 1 refers to different design metrics such as energy efficiency (EE), lifetime (LT), low latency (LL), fault tolerance (FT), security (S), quality of service (Q), and reliability (R), which are considered with the level of importance [low (L), medium (M), and high (H)] for different WSN applications.

This chapter’s main objective is the better understanding of data collection protocol with respect to network lifetime, energy conservation, fault tolerance, and low latency. In addition to this, understanding of some existing techniques such as multi-hop, clustering, duty cycling, aggregation, directional antennas, network coding, sink mobility, and cross-layer solutions for achieving these parameters.

2. Data collection

For sensing the data from the environment and transferring to the BS, the sensor nodes are deployed at specific locations. The data collection’s main goal is accuracy of sensing and transmitting the data to BS without any information loss and delay.
Transmitting of sensed data to BS is either by data dissemination (data diffusion) or data gathering (data delivery) [8]. Data/queries (network setup/management and/or control collection commands) propagation throughout the network is done in the data dissemination stage. Low latency is the main issue for disseminating data/queries to BS.

Data delivery or data gathering is the forwarding of sensed data to the BS. The main aim of data gathering is to maximize the number of rounds of data transferring toward BS before the network died. This will be achieved by minimizing energy consumption and delay for each transmission.

Single-hop or multi-hop is the basic communication technique between source sensor node and BS in data gathering. Sensed data are forwarded directly to BS in the single-hop communication. In multi-hop [9], the sensed data are forwarded to the base station with the help of intermediate sensor nodes. In multi-hop routing, energy conservation, route discovery, QoS, and low latency are the major issues. Introducing mobility in sink nodes, called mobile sinks or mobile collectors [10] is also a single-hop communication. In this network, mobile sink nodes move along a trajectory path to access the data from all source sensor nodes in a single-hop fashion. The trajectory path identification is the important step in this single-hop communication to cover all the nodes throughout the network. Energy conservation and mobility are the major issues in mobility-based single-hop data transmission.

2.1 Taxonomy of data collection protocols

Different classification of data collection routing protocols [6, 11–15] are proposed in recent years by researchers. Figure 1 shows the different classifications of data collection routing protocols.

Network architecture-based classification was presented by Akkaya et al. [6] in 2005. According to Akkaya et al., routing protocols are classified as data-centric, hierarchical, and location-based protocols. Sink disseminating the queries in network to get the sensor data from sensor nodes is the work of data-centric protocols. In cluster- or hierarchical-based protocols, network of nodes is divided into clusters and each cluster is managed by the cluster head (CH). Each CH will receive the sensed data from the corresponding cluster member and forward it to the BS. Aggregation techniques can be used by the CH to save energy while forwarding to BS. Geographic- or location-based protocols are considering the position information of sensor nodes for routing.

Multipath, query-based, negotiation-based, quality of service (QoS)-based, and coherent-based protocols are the classification of routing protocols as given by Karaki et al. [11]. In multipath routing, multiple paths are selected for achieving a variety of benefits such as reliability, fault tolerance, and increased bandwidth. Data
acquisition is done by the sink node with the help of query dissemination in query-based routing. All sensor nodes are going to store the data based on the interest of nodes. Then the data are forwarded to the destination only if the sensed or received node data match with the received queries. Data descriptors are used by negotiation-based protocols for reducing redundant data relays through negotiation. QoS-based protocols mainly consider QoS metrics such as delay, throughput, bandwidth, etc., when routing the data to the base station. In coherent routing, the sensed data is transferred directly to the aggregate node. Whereas in noncoherent routing, node data processing is done locally and then is transferred to neighbor nodes. In addition, routing protocols are classified into proactive, reactive, and hybrid protocols depending on path establishment between the source and destination.

Continuous, event-driven, observer-initiated, and hybrid-based on application interest are the different classifications given by Tilak et al. [12] in 2002. The sensor nodes transfer their sensed data at a prespecified rate to the server in the continuous model. Only when an event occurs, the sensor nodes forward data to base station in the event-driven data model. In the observer-initiated model, the observer will give an explicit request, then only the corresponding sensor nodes respond with the results. The combination of above three approaches will be called as hybrid protocols.

Based on data communication functionalities of routing protocols, Kai Han et al. [31], in 2013, classified the routing protocols into unicast, anycast, broadcast, multicast, and converge-cast. One-to-one association between sensor nodes is used in unicast routing. For forwarding the sensed data, unicast routing is using one neighboring node as a relay node. In anycast routing, nodes transfer the sensed data to a potential receiver node of a group. Multicast routing is transferring the data to a selected number of neighbor nodes simultaneously in a single transmission. Broadcast routing uses a one-to-many association; in a single transmission, sensor nodes transfer the data to their all neighbor nodes simultaneously. The data are aggregated at relay nodes and forwarded toward the base station in the converge-cast mechanism. Information exchanges will be done between the pair of sensor nodes in unicast/anycast. Whereas, multicast/broadcast is required for disseminating commands to sensor nodes, and converge-cast uses to collect the data from sensor nodes.

Routing protocols are classified as classical and swarm intelligence-based protocols by A.M. Zungeru et al. [14]. Further, each protocol is categorized into data-centric, hierarchical, location-based, network flow, and quality of service (QoS) awareness. In addition, they divided the routing protocols into proactive, reactive, and hybrid, depending on the path establishment between the source and destination.

The energy-efficient routing protocols are classified into network structure, communication model, topology-based, and reliable routing, as presented by Pantazis et al. [15]. Network structure routing protocols are classified into flat and hierarchical protocols. Communication model routing protocols can be divided into coherent or query-based and negotiation-based or noncoherent-based protocols. Mobile agent-based or location-based routing protocols are under the category of topology-based routing protocols. Reliable routing protocols are classified as multipath-based or QoS-based.

In addition to the above, some other literature [16–20] also presented different classifications of routing protocol. However, Figure 1 represents the overall classification of routing protocols in WSN.

3. Major design issues and techniques for data collection

In this section, some common design issues for data collection, such as energy, lifetime, latency, and fault tolerance are discussed. The techniques such as
clustering, aggregation, network coding, duty cycling, directional antennas, sink mobility, and cross-layer solutions which are used to achieve efficient data collection routing protocols are also presented.

3.1 Design issues in data collection

3.1.1 Energy and lifetime

Managing energy of the sensor nodes is the primary concern in WSN because it is the critical constraint of the sensor nodes. Saving of the node energy increases the network lifetime. Sensor node depletes much energy in two significant operations such as environment sensing and communicating sensed data to the BS. Energy consumption is stable for sensing operation because it depends on the sampling rate and does not depend on the other factors such as the topology of network or the location of the sensors. While, data forwarding process depends on them. Hence, energy conservation is feasible by designing an effective data forwarding process. Network lifetime [21] is defined as the period from the starting of the WSN operation to the time when any or a given percentage of sensor nodes die. Hence, the major objective of the data collection protocol is to gather the data with the maximum number of rounds within the lifetime of the network. The data gathering is the vital factor which considers energy saving as well as lifetime. In literature [4, 22], the authors have presented energy-efficient techniques for data collection. Rault et al. [4] have reviewed the energy-saving techniques and its classification such as radio optimization, data reduction, sleep/wake-up schemes, energy-efficient routing, and battery repletion. Anastasi et al. [22] in 2009 discussed directions for energy conservation in WSNs and presented the taxonomy of energy conservation techniques such as duty cycling, data driven, and mobility-based routing.

3.1.2 Latency

Latency is the period from the time unit that the data generation at the sensor node started to the time unit that data reception was completed at the base station. It is one of the main concerns for time significant applications such as military and medical health-care monitoring. Attaining low latency is a vital concern because of the following reasons:

1. Due to limited constraints of sensor nodes which are more prone to failure.

2. Collisions and network traffic will be increased due to the broadcast nature of radio channel.

3. Same kind of data will be sensed by densely deployed sensors and transfer to BS will increase the network traffic and exhaust the communication bandwidth.

To deal with the above issues, there is a need for low-latency protocols. Literature [23, 24] presents recent survey works on low-latency routing protocols. Srivathsan and Iyengar [23] have reviewed some key mechanisms to reduce the latency in single-hop and multi-hop wireless sensor networks; such mechanisms are sampling time, propagation time, processing time, scheduling, use of directional antennas, MAC protocols, sleep/wake-up cycles, predictions, use of dual-frequency radios, etc.
A review on energy-efficient and low-latency routing protocols for WSNs without dominating the other design factors is presented by Bagyalakshmi et al. [24].

### 3.1.3 Fault tolerance

Fault tolerance [25] enhances the availability, reliability, and dependability of the system by ensuring the usage availability of the system without any disruption in the presence of faults. In WSN, fault tolerance is also a demanding issue due to the sensor nodes more vulnerable to failure because of energy depletions, desynchronization, communication link errors, etc., which are provoked owing to hardware and software failures, environmental conditions, etc. Hence, fault management in WSN must be administered with additional care. Initial review works on fault-tolerant routing schemes are present in literature [21, 25–28]. Yu et al. [26] have explained issues in the fault management of WSN. Three phases called fault diagnosis, fault detection, and fault recovery for supervising faults have been proposed. In fault detection phase, an unexpected failure should be identified by the system. Literature [26–28] explains various fault detection techniques. In fault diagnosis phase, comprehensive description or model has been determined to distinguish various faults in WSNs [21] or fault recovery action. In the fault recovery phase, the sensor network is redesigned from failures or fault nodes to enhance the network performance. Fault recovery techniques have been dealt by literature [25].

### 3.2 Major techniques used for data collection design issues

The major techniques utilized for attaining energy saving, low latency, long lifetime, and fault tolerance in WSNs are discussed in this section.

#### 3.2.1 Cluster architecture

Cluster-based architecture is a foremost technique for effective energy conservation. In this mechanism, the network is partitioned into clusters, where the cluster head (CH) is a leader to manage the members of each cluster. Every member sensor node transmits the sensed data to their corresponding CH; then, CHs communicate the collected data to the BS. This technique avoids flooding, routing loops, and multiple routes; hence, reduced network traffic and low latency are attained. The major advantage of cluster-based architecture is that it needs less transmission power because of small communication ranges within the cluster. The CH uses the fusion mechanism to minimize the size of the transmission data. CH selection is performed in a rotation basis to balance the energy consumption in the network and improve the network lifetime. However, in cluster-based routing protocols, cluster head selection plays a critical role. Further, clustering algorithms do not consider the location of the base station, which creates a hot spot problem in multi-hop wireless sensor networks.

#### 3.2.2 Data aggregation

Data aggregation is one of the significant methods applied to aggregate the raw data evolved from multiple sources. In data aggregation schemes, nodes receive the data, reduce the amount of data by employing data aggregation techniques, and then transmit the data to the BS. The average or minimum amount of received data are merely forwarded by the received node. This reduces the network traffic and hence low latency is achieved. However, the base station (sink) cannot ensure the
accuracy of the aggregated data that have been received by it and also cannot restore the data.

3.2.3 Network coding

Network coding is the same as the aggregation technique. In this technique, the nodes collect the data from neighbor nodes and combine them together by applying mathematical operations; then it transmits data to the BS. This technique improves the network throughput, reliability, energy efficiency, and scalability; it is also resilient to attacks and eavesdropping. Network traffic in broadcast scenarios can be reduced by combining several packets as a single packet rather than sending separate packets.

3.2.4 Duty cycling

For energy conservation, duty cycling is one of the important techniques in WSNs. In duty cycling, the radio transceiver mode of sensor node is changing between active and sleep. This technique requires cooperative coordination between nodes for communication. Nodes want to communicate with each other and the nodes will shift from sleep mode to wake-up mode. A node must wait for its neighbor nodes to awake for communication. Sleep latency is increased due to this. Multi-hop broadcasting is complex in this technique because all the neighboring nodes are not active at the same time.

3.2.5 Directional antennas

Transmitting or receiving signals with one or more directions at a time with greater power is done with directional antennas. This technique improves the performance with respect to throughput by increasing the transmission range. With the help of directional antennas, bandwidth reusability is also possible. However, transmission power calculations and optimal antenna pattern selection overhead is more in these directional antennas. Also, directional antennas are more exposed to hidden and exposed terminal problems.

3.2.6 Sink mobility

Sink mobility is one of the energy-efficient technique, where mobility is introduced with sink nodes. The mobile sink nodes collect the data from sensor nodes with single-hop while moving in a specified path and then forward the same to the BS. This scheme reduces the workload of nodes which are placed nearer to the sink nodes and it increases the network lifetime. With the help of sink mobility, so many sparse networks can be connected and communicated which in turn provides scalability of the network. Reliability will be improved because of single-hop communication between the mobile sink and sensor nodes. However, trajectory path maintenance is a critical part of sink node while moving. Mobile collector needs a proper synchronization mechanism with sensor nodes, otherwise this causes packet loss while data gathering.

3.2.7 Cross-layered approach

When compared to layered approaches, cross-layered approach in WSN is energy efficient. The protocol stack is considered as a single system instead of individual layers in the cross-layered approach. For interaction among the protocol
layers, state information of the protocols is shared among all layers. Cross-layered protocol implementations significantly affect the system efficiency with respect to the energy and lifetime.

4. Existing routing techniques

In WSN, so many techniques are proposed to achieve energy efficiency, longer lifetime, fault tolerance. Low latency by different researchers are briefly explained in this section. Most of these solutions are designed based on different techniques such as clustering, network coding, duty cycling, aggregation, directional antennas, sink mobility, and cross-layer solutions.

Low-energy adaptive clustering hierarchy (LEACH) routing strategy was proposed by Heinzelman et al. [29]. It is a cluster-based routing algorithm to decrease energy consumption and improve the network lifetime. In this protocol, the network is divided into clusters; each cluster contains a set of CMs and a leader called CH. The CMs send the data to its respective CH; CHs communicate the collected data to the BS and are elected in a random and distributed manner. Subsequently, LEACH was altered to LEACH-C [30], a centralized approach. The process of CH selection is performed based on the residual energy of the sensor nodes. However, due to dynamic cluster formation, the distance between CH and BS is faraway and some of the cluster nodes are also faraway from the CHs; it increases the communication cost. Later, a lot of modified LEACH protocols have been proposed to enhance the network lifetime and have been reviewed in [17].

LEACH protocol has been improved as power-efficient gathering in information systems (PEGASIS) [31], a multi-hop chain-based protocol, where every node aids in transmitting and/or receiving the data from its neighbor node by forming the chain. The collected data are aggregated and carried from node to node. One of the nodes in the chain is selected as a leader; the leader node transfers data to the BS. PEGASIS performs better than LEACH by minimizing the number of transmissions from sensor nodes to BS and clustering overhead. However, data transmission delay is higher due to the large chain length.

Threshold-sensitive energy-efficient sensor network protocol (TEEN) [32] is a homogenous reactive routing protocol. In this approach, the process of CH selection is performed similar to LEACH; the data transmission varies from LEACH. The workings of TEEN are based on the thresholds, namely, Hard threshold ($H_T$) and soft threshold ($S_T$). However, the CH selection process is random and the size of the clusters is unequal; it causes an unbalanced energy consumption among the clusters. Network throughput is also decreased due to the threshold mechanism.

Hybrid energy-efficient distributed (HEED) protocol [33] has been proposed by Younis and Fahmy. It is a homogenous cluster-based routing protocol; CH selection is accomplished based on the probability function of residual energy and node degree. Later, HEED protocol is extended as the heterogeneous HEED to manage the routing in the heterogeneous network field. This protocol utilizes fuzzy logic model for the CH selection process; the parameters considered in the fuzzy logic model are node degree, distance, and remaining energy. Finally, direct data transmission is carried out between the CM and CH and between the CH and BS.

Qing et al. [34] have presented distributed energy-efficient clustering scheme (DEEC), a heterogeneous data collection protocol. The sensor nodes possess varied energy levels. The selection of CHs is done based on the probability ratio between the residual energy of the nodes and average energy of the whole network. The possibility of evolving a CH is higher for the nodes which possess more residual energy.
energy. However, the probabilistic CH selection process prompts unequal clusters which leads to more energy dissipation.

Periodic, event-driven, and query-based protocol (PEQ) and its variation, CPEQ, were proposed by Boukerche et al. [35] in 2006. PEQ is designed for achieving the following: low latency, high reliability, and broken path reconfiguration. CPEQ is a cluster-based routing protocol. The publish/subscribe mechanism is used to broadcast requests throughout the network.

Genetic algorithm-based clustering approach (LEACH-GA) was introduced in literature [36] to predict the optimal probability for electing an optimal number of CHs. This approach improved the network lifetime by achieving energy-efficient clustering.

Artificial bee colony (ABC)-based algorithm [37] has been proposed, where the CH selection is performed by adopting the ABC algorithm. ABC algorithm improves the clustering process by employing efficient and fast search feature to select the CHs. Both cluster members to CH, and CH to BS communication is performed by direct data communication. However, this protocol does not consider the coverage of the CH and it prompts more energy dissipation.

Ant colony algorithm for data aggregation (DAACA) has been introduced by Chi Lin et al. [38]. This approach comprises of three phases: initialization, packets transmissions, and operations on pheromones. In the transmission phase, the next hop is dynamically selected by determining the number of pheromones of neighbor nodes and the residual energy. Pheromones adjustments are accomplished for every specified number of rounds of data transmissions. Besides, various pheromones’ adjustment strategies such as basic-DAACA, elitist strategy-based DAACA (ES-DAACA), maximum- and minimum-based DAACA (MM-DAACA), and ant colony system-based DAACA (ACS-DAACA) are utilized to enhance the network lifetime. However, duplication packets are transmitted from sink nodes to initialize the network, which causes higher energy depletion in the network.

Lusheng Miao et al. [39] have introduced network coding to resolve the issues in gradient-based routing (GBR) scheme, such as broadcasting of interest messages by sink node which prompts duplication of packets, which causes more energy dissipation, and point-to-point message delivery forces more data retransmissions due to the unstable network environment in WSNs. The authors have proposed network coding for GBR (GBR-NC) to implement energy-efficient broadcasting algorithm which reduces network traffic. Further, the authors have presented two competing algorithms such as GBRC and auto-adaptable GBR-C to minimize the data retransmissions.

In 2012, Rashmi Ranjan Rout et al. [40] proposed an energy-efficient triangular (regular) deployment strategy with directional antenna (ETDDA), where 2-connectivity pattern has been utilized. This pattern is accomplished by aligning the directional antenna beam of a sensor node in a specified direction toward the sink. Data forwarding depends on network coding for many-to-one traffic flow from sensor nodes to sink. The proposed approach ensures energy efficiency, robustness, and better connectivity in communicating data to the sink.

Ming Ma et al. [41] have put forward a mobility-based data-gathering mechanism for WSNs. A mobile data collector (M-collector), perhaps a mobile robot or a vehicle, is implemented with a transceiver and battery. The M-collector travels through a specific path and determines the sensor nodes, which comes within its communication range while traversing. Then, it collects the data from the sensor nodes in the single-hop communication and forward the data to the base station without delays. Hence, this mechanism improves the lifetime of the sensor nodes. The authors have primarily focused to reduce the length of each data-gathering tour called as single-hop data-gathering problem (SHDGP).
Roja Chandanala et al. [42] have presented a mechanism to preserve energy in flood-based WSNs by applying two techniques: network coding and duty cycling. Initially, the authors have proposed DutyCode, a cross-layer technique, where Random Low Power Listening MAC protocol was devised to implement packet streaming. The authors have applied flexible intervals for randomizing sleep cycles. Further, an enhanced coding scheme was proposed, which selects appropriate network coding schemes for nodes to remove redundant packet transmissions.

Meikang Qiu et al. [43] have introduced informer homed routing (IHR), which is a novel energy-aware cluster-based fault-tolerance mechanism for WSN. IHR is the foremost variant of dual homed routing (DHR) fault-tolerance mechanism. In this mechanism, each sensor node is attached with two cluster heads called primary cluster head (PCH) and backup cluster head (BCH). Sensor nodes deliver the data to PCH rather than sending simultaneously to both PCH and BCH. In each round, BCH probes the PCH to identify whether the PCH is active or not using the beacon message. In three continuous rounds, if BCH cannot receive any beacon message from PCH, then BCH will declare that the PCH has failed and it informs to sensor nodes to transmit data to BCH. Hence, IHR provides an energy-efficient fault-tolerance mechanism to prolong the lifetime of the network. However, cluster head selection process is containing more overhead.

A novel evolutionary approach for load-balanced clustering problem is presented in literature [44]. CH (gateway) formation is performed using a novel genetic algorithm. This algorithm differs from the traditional GA in the initial population and mutation phase. This approach balances the load among the gateways and it is energy efficient. However, sensor nodes that are not reachable to any gateway are left out from communication. Later, they extended a differential evolution-based approach [45] used for clustering the nodes with gateways (CHs) in a load-balanced way to ensure load balancing among the gateways and energy efficiency. But, this approach used single-hop communication between the gateway to BS and hence it may not be suitable for long-distance communication.

Flow partitioned unequal clustering (FPUC) algorithm has been proposed by Jian Peng et al. [46] to attain an enhanced network lifetime and coverage. FPUC has two phases: clustering and flow partition routing. In the clustering phase, cluster head is decided based on the higher residual energy and larger overlapping degree of sensor nodes. In the flow partition routing phase, cluster head collects the data from the member nodes and aggregates the data into a single packet; then it forwards the data to the sink through gateway nodes depending on residual energy. The flow-partitioned routing phase has two subphases: dataflow partitioning phase and relaying phase. In the dataflow partitioning phase, the cluster head segments the dataflow into various smaller packets and then delivers these packets to its gateway nodes. In the relaying phase, gateways communicate the received data to the next hop with minimum cost.

An energy-efficient adaptive data aggregation strategy using network coding (ADANC) to attain improved energy efficiency in a cluster based duty-cycled WSN has been introduced by Rashmi Ranjan Rout et al. [47]. Network coding minimizes the network traffic inside a cluster and duty cycling scheme has been used in the cluster network to prolong network lifetime.

Dariush Ebrahimi and Chadi Assi [48] have presented a new compressive data gathering method. This method utilizes compressive sensing (CS) and random projection techniques to enhance the lifetime of large WSNs. The authors preferred the method to equally distribute the energy throughout the network rather than decreasing the overall network energy consumption. In the proposed data-gathering method, minimum spanning tree projection (MSTP) has been adopted. MSTP creates several minimum spanning trees (MSTs) and each root node of the tree
aggregates sensed data from the sensor nodes using compressive sensing. A random projection root node with compressive data-gathering aids to achieve a balanced energy consumption all over the network. Besides, eMSTP has been introduced which is the extended version of MSTP; the sink node in the eMSTP behaves like a root node for all MST.

Ahmad et al. [49] proposed a protocol called Away Cluster Heads with Adaptive Clustering Habit (ACH^2) and this mechanism has been utilized for enhancing network lifetime. However, global node information is required for communicating data and the size of the clusters is also unequal. As the node distribution among the clusters is unequal, this approach prompts to variation in energy depletion ratio among clusters in the network.

A genetic algorithm-based approach [50] has been applied for binding the sensor nodes to the sink nodes, considering the balanced load among the sink nodes. The authors have presented a fitness function which takes into account the communication cost between the sensor node and sink node and the processing cost of the sink node. This approach dealt with the nodes which do not have any sink node in their communication range.

In 2015, energy-aware routing (ERA) [51] has been proposed, where the residual energy of the CHs and the intra-cluster distance are the parameters taken into account for the process CH selection. However, the parameters such as the optimal number of CHs, network density, and cluster coverage are not considered in the CH selection process; hence this causes uneven energy consumption in every cluster.

A GSA-based approach titled GSA-based energy-efficient clustering (GSA-EEC) was presented by literature [52]. For the fitness value calculation, the parameters considered are the distance between the sensor nodes and gateways, the distance between gateways and sink, and residual energy of gateways. This approach improves the network lifetime and total energy consumption. Further, they introduced a routing strategy titled gravitational search algorithm-based multi-sink placement (GSA-MSP) for placing multiple sinks on the sensor network [53].

Priority-based WSN clustering of multiple sink scenario using artificial bee colony [54] has been proposed. The fitness function in this approach considers the energy of the sink node and the sensor node, the distance between the sensor node to the sink node, and the priority of each sink.

PSO-based approach for energy-efficient routing and clustering has been proposed in literature [55]. Routing path between the gateway to BS is determined using the PSO technique. This approach provides energy-efficient routing and energy-balanced clustering. This approach is fault tolerant when CHs failed. But, nodes that are not reachable to any gateway are left out from communication.

Gravitational search algorithm for cluster head selection and routing (GSA-CHSR) [56] has been proposed. The authors have used GSA algorithm for deciding the optimal number of CH nodes and finding the optimal route between CH and BS. This approach improves performance parameters such as network lifetime, residual energy, and the number of packets received at BS. However, this approach incurs clustering overhead for selecting the optimal set of CHs.

Guravaiah and Leela Velusamy [57] proposed a routing protocol titled hybrid cluster communication using RFD (HCCRFD) based on clustering using river formation dynamics-based multi-hop routing protocol (RFDMRP) [58]. This protocol increases the network lifetime. However, load balancing among CHs is not considered and clustering overhead exists due to periodic CH selection. Further, the authors have proposed a balanced energy and adaptive cluster head selection algorithm (BEACH) [59]. They considered the parameters such as degree of the node, remaining energy of the node, the distance from BS to the sensor node, and the average transmission distance to its neighbors for achieving the load-balanced clustering.
<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Algorithm</th>
<th>Techniques used</th>
<th>Metrics</th>
<th>Drawbacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PEGASIS [31]</td>
<td>Chain construction</td>
<td>Lifetime</td>
<td>Network throughput decreased</td>
</tr>
<tr>
<td>2</td>
<td>LEACH [30]</td>
<td>Clustering</td>
<td>Lifetime, scalability</td>
<td>Not considering RE for CH selection, unbalanced energy consumption</td>
</tr>
<tr>
<td>3</td>
<td>TEEN [32]</td>
<td>Clustering</td>
<td>Lifetime</td>
<td>Same as LEACH, network throughput decreased</td>
</tr>
<tr>
<td>4</td>
<td>HEED [33]</td>
<td>Clustering</td>
<td>Lifetime</td>
<td>Direct transmission, heterogeneity is not considered</td>
</tr>
<tr>
<td>5</td>
<td>DEEC [34]</td>
<td>Clustering</td>
<td>Lifetime, scalability</td>
<td>Direct transmission, unequal size of clusters, unbalanced energy consumption</td>
</tr>
<tr>
<td>6</td>
<td>PEQ and CPEQ [35]</td>
<td>Clustering and publish/subscribe mechanism</td>
<td>Fault tolerance, low latency, and energy</td>
<td>Traffic overhead</td>
</tr>
<tr>
<td>7</td>
<td>DAACA [38]</td>
<td>Clustering, ACA</td>
<td>Energy, network lifetime</td>
<td>Bottleneck problem nearer to sink node, overhead in pheromones calculation at each round</td>
</tr>
<tr>
<td>9</td>
<td>ETD-DA [40]</td>
<td>Directional antennas, network coding, and multi-hop</td>
<td>Energy, throughput, and low latency</td>
<td>Overhead in optimal antenna pattern and transmission power calculations</td>
</tr>
<tr>
<td>10</td>
<td>SHDGP [41]</td>
<td>Mobile collectors and single-hop</td>
<td>Energy, low latency, scalability, and throughput</td>
<td>High control overhead to maintain the trajectory path, packet loss due to speed of data collector</td>
</tr>
<tr>
<td>11</td>
<td>DutyCode and ECS [42]</td>
<td>Network coding, duty cycling, and multi-hop</td>
<td>Energy</td>
<td>Transition between active and sleep states overhead</td>
</tr>
<tr>
<td>12</td>
<td>IHR [43]</td>
<td>Clustering and multi-hop</td>
<td>Fault tolerance and energy</td>
<td>Node unable to find CH, leads to reliability problems</td>
</tr>
<tr>
<td>13</td>
<td>Novel evolutionary approach [44]</td>
<td>Clustering, genetic algorithm</td>
<td>Energy efficiency</td>
<td>Single-hop communication between the CH to BS</td>
</tr>
<tr>
<td>14</td>
<td>DE-based clustering algorithm [45]</td>
<td>Clustering using differential evolution</td>
<td>Energy efficiency</td>
<td>Single-hop communication between the CH and BS</td>
</tr>
<tr>
<td>15</td>
<td>FPUC [46]</td>
<td>Clustering, data aggregation, and multi-hop</td>
<td>Energy efficiency, lifetime</td>
<td>CH selection overhead</td>
</tr>
<tr>
<td>16</td>
<td>ADANC [47]</td>
<td>Clustering, network coding, and duty cycling</td>
<td>Energy, low latency, and lifecycle</td>
<td>Cluster maintenance overhead</td>
</tr>
<tr>
<td>17</td>
<td>MSTP [48]</td>
<td>Data aggregation using compressive sensing</td>
<td>Energy, network lifetime</td>
<td>Computational overhead in MST calculations</td>
</tr>
</tbody>
</table>
An approach called LEACH-PSO [60] has been proposed for improving the network lifetime by selecting an optimum number of CHs in every round. In this work, the particle swarm optimization method is integrated with LEACH for forming the clusters.

Energy-efficient CH-based GSA (GSA-EC) [61] for finding an optimal set of CHs using GSA has been proposed. To balance the energy consumption, one-hop clusters are formed using an optimal set of CHs. The authors have also proposed the hybrid approach of PSO and GSA. This approach increases network lifetime and network stability. However, this approach also incurs clustering overhead for selecting the optimal set of CHs. Later, Kavitha et al. [62] used GSA for assigning sensor nodes to an appropriate cluster head (CH) in a load-balanced way such that it reduces the energy consumption and hence enhances the lifetime of a network.

Integrated clustering and routing protocol using cuckoo and harmony search has been proposed in literature [63]. This approach has adopted the cuckoo search algorithm for CH selection. Residual energy, degree of a node, intra-cluster distance, and coverage ratio are the parameters for developing fitness function used in CH selection. The harmony search algorithm has been employed for routing from

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Algorithm</th>
<th>Techniques used</th>
<th>Metrics</th>
<th>Drawbacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>ACH² [49]</td>
<td>Clustering</td>
<td>Lifetime, throughput</td>
<td>Global node information for data transmission, cluster head selection overhead</td>
</tr>
<tr>
<td>19</td>
<td>GA-based approach [50]</td>
<td>Clustering using genetic algorithm</td>
<td>Energy efficiency, lifetime</td>
<td>Single-hop communication between sink and BS</td>
</tr>
<tr>
<td>20</td>
<td>Energy-aware routing (ERA) [51]</td>
<td>Clustering</td>
<td>Energy efficiency, lifetime</td>
<td>Optimum number of CHs is not considered</td>
</tr>
<tr>
<td>21</td>
<td>GSA-EEC [52]</td>
<td>Clustering, GSA</td>
<td>Energy efficiency</td>
<td>Load balancing among CHs not considered</td>
</tr>
<tr>
<td>22</td>
<td>PSO-based routing [55]</td>
<td>Clustering (gateways) using PSO</td>
<td>Energy efficiency</td>
<td>Nodes that are not reachable to any gateway are not considered</td>
</tr>
<tr>
<td>23</td>
<td>GSA-CHSR [56]</td>
<td>Clustering using multi-hop GSA</td>
<td>Energy efficiency</td>
<td>Clustering overhead</td>
</tr>
<tr>
<td>24</td>
<td>HCCRFD [57]</td>
<td>Clustering using LEACH, RFD</td>
<td>Energy efficiency</td>
<td>No load-balanced clustering</td>
</tr>
<tr>
<td>25</td>
<td>BEACH [59]</td>
<td>Clustering, RDF</td>
<td>Energy efficiency</td>
<td>CH selection overhead</td>
</tr>
<tr>
<td>26</td>
<td>GSA-EC [61]</td>
<td>GSA, multi-hop</td>
<td>Network lifetime</td>
<td>Clustering overhead</td>
</tr>
<tr>
<td>27</td>
<td>Cuckoo and harmony search-based routing [63]</td>
<td>Cuckoo search algorithm</td>
<td>Energy efficiency</td>
<td>Load balance among CHs not considered</td>
</tr>
<tr>
<td>28</td>
<td>MLBC [64]</td>
<td>MOPSO, multi-hop, spanning tree</td>
<td>Energy efficiency, reliable</td>
<td>Nodes that are not reachable to any CH are not considered</td>
</tr>
<tr>
<td>29</td>
<td>Energy-efficient and delay-less routing [65]</td>
<td>FCR algorithm</td>
<td>Energy efficiency</td>
<td>Energy balancing is not ensured</td>
</tr>
</tbody>
</table>

Table 2. Existing protocols for data collection.
CH to BS. It is energy efficient and balances the energy consumption of the network. Further, it minimizes the un-cluster nodes, that is, nodes that are not within the communication range of any CH are minimized. But, load balancing among CHs is not considered.

Multi-objective load-balancing clustering technique (MLBC) [64] has been proposed for clustering in WSN by adopting multi-objective PSO (MOPSO) strategy which is used for CH selection. The shortest-path tree (SPT) for loop-free routing is created using Dijkstra’s algorithm. It is energy efficient and reliable. But, the nodes that are not reachable to any CH are not considered.

In energy-efficient and delay-less routing [65], CH selection is performed using firefly with cyclic randomization (FCR) algorithm. This approach reduces transmission delay in the network. But, this approach has not considered energy balancing.

Overall comparison of above routing protocols are shown in Table 2 with the techniques used, metrics considered, and drawbacks of each solution.

5. Future directions

Overall, the above discussed techniques’ main objective is energy-efficient data gathering and is concentrated on the following issues:

• Duplication of data generation and forwarding
• Congestion or data storm problem nearer to the base station
• Selection of multi-hop routing path
• Operations to perform data aggregation
• Selection of cluster head

However, we need to concentrate on the following future directions for proposing new routing techniques:

• Almost all protocols require location information for routing. Location finding can be done using localization or GPS techniques, which are dependent on energy consumption. Finding of sensor location with less consumption of energy is an issue.

• Most of the multi-hop routing protocols suffer from overheads and delay due to path setup and relay nodes. Also, formation of loops in aggregate tree generation increases the energy consumption.

• Most of the literature failed in energy calculations at the time of CH selection in cluster-based routing protocols.

• Uneven distribution of cluster heads will generate unequal-sized clusters, unbalanced energy consumption between cluster members, and CH coverage problem.

• The size (with respect to area and number of members) inequality among the clusters leads to network coverage problem due to limited communication
• The sizes of the clusters formed in the existing protocols are not equal. This leads to unbalanced energy consumption among the clusters.

• Density of network was not considered as a parameter in CH selection process. This impacts the formation of unequal sized clusters and leads to uneven distribution of load to CH.

• Uneven distribution of load on CH and the intra- and inter-communication path length is more.

• Security is the major parameter need to be considered in military applications. Considering security, energy efficiency is still challenging issues.

• In recent years, more popularity gain is deterministic rather than probabilistic-based clustering due to reliability. However, CH selection and other computational complexity are still a challenging area.

• Heterogeneous network in WSN is also an important problem due to different communication and processing capabilities.

6. Conclusions

In this chapter, classification of data collection routing protocols in WSN has been thoroughly discussed. Various techniques such as clustering, duty cycling, aggregation, network coding, sink mobility, and cross-layered solutions, and directional antennas have been utilized by data collection routing protocols for attaining long lifetime, energy efficiency, fault tolerance, and low latency. These techniques are reviewed briefly in this chapter. Finally, this chapter demonstrates a paramount comparison among the existing approaches applicable on data collection process in WSN. Future directions of routing protocols are presented at the end of this chapter.
Author details
Koppala Guravaiah\textsuperscript{1,\dagger}, Arumugam Kavitha\textsuperscript{2,\dagger} and Rengaraj Leela Velusamy\textsuperscript{2,\dagger}

1 Indian Institute of Information Technology, Kottayam, Kerala, India

2 National Institute of Technology, Tiruchirappalli, Tamil Nadu, India

*Address all correspondence to: kguravaiah@iitiottayam.ac.in

\textdagger All authors have contributed equally.
References


[34] Qing L, Zhu Q, Wang M. Design of a distributed energy-efficient clustering


[50] Safa H, Moussa M, Artail H. An energy efficient genetic algorithm based approach for sensor-to-sink binding in


