Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 3521 - 3539

Received 15 December 2020; Accepted 05 January 2021.

Data Monitoring and Gathering in Sensor Networks for Healthcare Applications

Nitin Mittal

Deptt. of ECE, Chandigarh University, Mohali-140413, Punjab, India
E-mail address: mittal.nitin84@gmail.com,

Abstract: The use of sensor technologies for monitoring the environment for a variety of purposes such as defense, health, education, comfort, the environmental, traffic, security, and the like would be an integral part of life. In contrast with traditional wireless networks, these systems need close coupling with physical environments. The networks of wireless sensors therefore present dramatically different design, deployment and usage challenges. The sensors should be networked so that measured and tracked parameters can be transmitted and disseminated to certain collection sites where more information is managed for decision making. Multiple sensors give the end user a fault tolerance and better parameter monitoring capability, both spatial and temporal, and can provide useful inferences of the physical world. In this paper, a review of data aggregation is performed for medical and healthcare applications.

Keywords: Wireless sensor networks, data aggregation, network lifetime, energy efficiency.

INTRODUCTION

Recent progress has been made in micro sensor technology and the development of distributor wireless sensor networks with a low level of power in analog and digital technology. For hundreds of cheap nodes to be installed in physical environments, the sensor networks of the future are intended to collect valuable information in a stable and autonomous way (e.g. seismic, acoustic, health and surveillance data). A distributed sensor network is typically a self-organized system consisting of a number of sensor nodes which collaborate to measure different parameters and send the relevant data to an additional collector center.

These sensor nodes are used to process, compute, and network self-configure signals to achieve scalable, stable, and long-lived networks. Data dissemination in sensor networks is often usually conducted as a joint process, due to power and range constraints, during which sensors work together to acquire data from the various sections of the sensor network from the information basin [1]. Due to its energy-efficiency and scalability, collaboration between various sensor nodes is mainly accomplished by multi hop network architectures[2].
A major energy consumer in wireless sensor networks is radio communication. Because local computation is much cheaper than radio communication, supporting some in-network processing [3]to reduce data within the network, can provide significant energy savings.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 3521 - 3539

Received 15 December 2020; Accepted 05 January 2021.

Several research works have discussed the problems of developing efficient joint routing and data management techniques, as well as application-specific data aggregation processes, to support efficient data monitoring and gathering approaches in sensor networks. Furthermore, several issues associated with the data aggregation process with the specific objective of meeting the task requirements (i.e., quality of service [QoS] constrained data monitoring and gathering) have also been recognized and considered recently.

The rest of this paper is structured as follows. We first present some performance metrics of interest for the design of efficient data monitoring and gathering methods, followed by a discussion of the related traffic and data correlation models. We then discuss several issues associated with the sensor node deployment patterns for the implementation of effective data monitoring. We also describe several data gathering approaches by classifying them into many diverse categories according to different features and design dimensions of the wireless sensor networks, and present some indicative associated protocols. Finally, we present a relative qualitative comparison of the different data gathering approaches, highlighting their respective merits, drawbacks, and performance.

Performance Metrics

Before we describe and discuss the various data gathering and dissemination approaches, we first present some performance metrics of interest and their impacts on the design of efficient and effective data monitoring and gathering methods. The sensor network should dynamically adapt to the system and topology changes; at the same time, it needs to balance the trade-offs among the various performance metrics. Following is a summary of the main performance metrics that need to be considered by the data monitoring and gathering approaches in sensor networks.

Delay/latency: For time-crucial applications, the sensor nodes are required to complete the data monitoring and gathering task within a predefined and strict latency; any data received out-of-date is considered useless. It may be defined as the time required for some collected data to be transmitted back to the collection sites. In most cases, of practical interest are some statistics with respect to the delay (e.g., average delay, probability that delay is less than some given threshold), rather than single delay values.

Energy efficiency and network lifetime: The parameters used to evaluate the degree of energy efficiency include the average energy used to transmit a bit to/from the source to the collection center (J/bit), the total energy consumption/ dissipation over an operation time period (J/unit time), and the ratio of the energy consumed to transmit the data payload to that consumed to transmit the overhead. Closely related to the energy efficiency is network lifetime

Accuracy: The precise definition of the accuracy is determined by the specific application. For example, if the sensor network is used to monitor environmental variables, the observed signal specifications such as temporal resolution, spatial resolution, and range accuracy are the accuracy parameters of concern.

Sensing coverage: According to the deployment schemes, the sensing coverage can be divided into two categories: deterministic and stochastic coverage. The stochastic coverage refers to the scenario in which the sensor field is covered with sensors randomly distributed

Received 15 December 2020; Accepted 05 January 2021.

throughout the environment. The sensing coverage relies not only on the appropriate placement of the nodes but also on their sensing capabilities.

Throughput: Because the bandwidth in sensor networks is limited and high node density may produce large amounts of data, the end-to-end transmission throughput needs to be maximized, besides providing fairness and low complexity of implementation.

Additional parameters: Some additional simplified and specific metrics, such as bit-hop metric [4], the ratio of payload data to overhead of packets, and the probability of buffer overflow have also been used to indicate the degree of energy efficiency. [5]proposed and used the product of the energy and delay as a potential metric to capture the system performance on balancing the energy and delay cost for data gathering in sensor networks. Alternatively, weighted cost functions can be used to balance the trade-offs of the various performance metrics[6]. The cost function can be defined as the sum of products of the weights and the corresponding performance elements (e.g., delay, energy efficiency, accuracy,where the weights can be adjusted according to the end-user demands and requirements[7].

Traffic Modeling And Data Correlation

Sensor networks are typically more application specific than traditional communication networks that are designed to accommodate various types of traffic and applications. As a result, closely related to the data monitoring and gathering process for different applications and scenarios are the corresponding traffic models and data correlation models.

Traffic Models

Depending on the patterns of the measurement and information to be monitored and transmitted, the traffic can generally be divided into three main categories: deterministic traffic, event-driven or threshold-sensitive traffic, and response-to-inquiry traffic. For the first type, there is a steady traffic flow between the source node and the sink. In such cases, sensors usually generate deterministic and/or periodic traffic, in which each node transmits its data once every specific time interval. The event-driven corresponds to cases where transmission of information is triggered by an event (e.g., a monitored variable exceeds some threshold; the change of an attribute value exceeds some predefined threshold, etc.). The sensor networks that aim at object detection or system monitoring (e.g., road traffic monitoring, forest fire monitoring, etc.) usually present this type of traffic. The third class, response-to-inquiry traffic, corresponds to cases where the sensors respond to inquiries dispensed by an end user or observer, which may be targeted to a specific set of sensors and/or for a specific time interval[8].

In general, various applications may involve more than one traffic modes. For example, in a seismic monitoring sensor network, the observation value will be periodically sent back to the collection site to record and accumulate data at the normal operation state for further analysis and processing; however, when the seismic intensity (vibration amplitude) measured exceeds a specific threshold, the observation will be transmitted to the sink immediately and the end users may further issue inquiry-based traffic. The different traffic and load conditions in the network may impact the performance of the data gathering approach[9]. For instance, the ad hoc response-to-inquiry traffic results in two-way communication that involves the dissemination of both the inquiries and the collected data, whereas the other two modes mainly generate one-way communication flows. The data dissemination and gathering strategies should be selected taking into account the different traffic models. For deterministic traffic, the data gathering strategy can be optimized based on certain known patterns, whereas in principle the data gathering strategy should be able to adapt to the burst change of the traffic, in order to accommodate other traffic modes as well[10].

Data Correlation Modeling

In sensor networks, the data in the neighboring nodes are considered highly correlated because the observed objects in the same geographical location are usually strongly correlated. Furthermore, utilizing the appropriate spatial correlation models to generate synthetic data based on the availability of only a small amount of experimental data inputs allows for efficient and accurate testing and evaluation of the corresponding data monitoring and gathering approaches [11].

With respect to the signals emitted from certain sources, it is usually assumed that the signal magnitude of the event's effect at a distance d from the source is proportional to $1/d^{\alpha}$, where α is the propagation parameter. The value of depends on the type of sensing event (e.g., seismic vibration, sound, light, infrared signal, etc.) and the medium in which the signals travel/propagate (e.g., ground, air, water). If the observation has multiple sources, the signal event arriving at a node is usually obtained as some function of the multiple events' effect.

In principle, the readings at each sensor can be regarded as samples of a random variable. Therefore, statistical joint moments of two random variables can be used to summarize the corresponding correlation properties. Assuming that the observations at two nodes are represented by random variables X and Y, respectively, then the ij-th joint moment of X and Y is represented by $E(X^iY^j)$. For i= j= 1, the resulting moment E(XY) is called the correlation of X and Y, whereas the covariance of X and Y is defined as: COV (X,Y) $E[(X - E(X))(Y - E(Y))]$, where E[X] and E[Y] denote the expected values of random variables X and Y, respectively. Both of these statistics are among the most commonly used to characterize the correlation between two observation sets. For sensors using n-bit A/D converters, based on the fact that a reading can be represented as one of the 2^n possible values whereas the difference of the readings among nearby nodes may be represented by fewer than n bits, [12]proposed a differential encoding method that allows nodes to transmit fewer bits for each reading. A linear model is used to estimate the correlation between two sample data of two nodes and to further determine the number of bits with which to ask the sensors to encode their values. According to this model, the optimal scaling factor that can provide best estimation between two known reading sets X and Y was found to be $E[XY]/E[Y^2]$.

Sensor Deployment For Data Monitoring And Gathering

The overall purpose of an efficient data monitoring and gathering methodology is to provide accurate and timely information, and at the same time extend the network operation lifetime for as long as possible. The optimal data gathering strategy depends on the density of nodes, position of sink, task requirements, and amount of correlation among the sources of the data. Therefore, to develop effective data monitoring and gathering strategies, we first need to

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 3521 - 3539

Received 15 December 2020; Accepted 05 January 2021.

develop an appropriate deployment scheme, which includes determining the number of nodes, density, types of sensor nodes, and how to deploy the nodes (e.g., deterministic or random), to achieve the required objectives, such as sensing coverage, fault tolerance, measurement accuracy, and so forth.

The placement of sensors has significant effects on many factors associated with the data monitoring and gathering in sensor networks. The deployment planning process involves a detailed analysis of the environmental map and related data to determine the most appropriate placement of sensor nodes and maximize the sensor fi eld coverage. In general, the sensing coverage reflects how well an area is monitored or tracked by sensors. Because of the inherent uncertainty associated with the processes of monitoring and sensing, probabilistic modeling of sensor coverage is desired. After the sensors are deployed according to a predetermined pattern, the locations of the nodes can be adjusted or additional sensor nodes may be placed to maintain the required sensor coverage and/or the desired network connectivity as the sensor network evolves.

Various deployment approaches that may range from deterministic to random have been considered by taking into account different constraints and situations. Specifically, for deterministic deployment, each node is placed at a specific position (manually or using robots) to maximize the sensor coverage. This method is suitable for cases where the sensors are static or fixed and sufficient knowledge of the environment is available for pre-calculation of the optimal sensor positions. [13]have developed approaches for deployment in two- and three-dimensional grids by formulating the node placement problem as combinatorial optimization[14], [15] and coding theory problems, and solving them using integer linear programming.

Although this kind of well-controlled node placement can provide good coverage for a given scenario. Therefore, the alternative option—random deployment of sensor nodes, such as throwing sensor nodes from air vehicles into the target area—is often more practical and desirable. The aforementioned stochastic coverage; [16]proposed a centralized polynomial time algorithm for the computation of worst-case and best-case coverage for random deployment using Voronoi diagram and graph search algorithms. In order to achieve energy efficiency, several algorithms have been proposed to address how to adaptively place sensors into the sleep mode while still maintaining full coverage of the sensing fields.

In some situations, it is expected that only very limited prior knowledge of the possible targets—or even no knowledge of the corresponding terrain—would be available. In this case, the single-step deployment may produce inferior coverage because of the lack of environmental information. One alternative method is to deploy sensors sequentially (i.e., to scatter a subset of the sensors at each step) and the information cumulated from previous deployed sensor nodes will be used to determine the further deployment steps. [17]developed an incremental deployment approach of sensor nodes for target detection purposes. For instance, [18] proposed polynomial-time algorithms, in terms of the number of sensors, to determine whether every point in the sensor field is covered by at least k sensors.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 3521 - 3539

Received 15 December 2020; Accepted 05 January 2021.

It should be noted here that a large number of uncertainties may occur in the sensor deployment process. For instance, when dispensing sensors from air vehicles, the actual landing position is affected by many elements, including the trace of air vehicles, the wind, the terrain conditions, and other obstacles such as trees and buildings. [19]proposed a new approach to address the uncertainty problem. For deterministic deployment, the expected sensor coverage and the actual results may differ because of the change of environment as the network evolves. Therefore, as previously mentioned, further adjustment may be required after the initial deterministic or random deployment. This could be achieved either by moving the deployed nodes to their desired location or by placing supplementary sensor nodes to cover the blind spots. Because the position of mobile nodes is easy to adjust, they are usually used to implement self-deployment or to provide enhanced flexibility. Toward this direction, [20]have designed distributed movement-assisted self-deployment protocols for mobile sensors.

Another important design challenge of a sensor network is the issue of self-organization. Self-organization is a critical attribute needed to achieve the wide use and applicability of distributed sensor networks. Consequently, once the sensors are deployed, they have to form networks in an autonomous matter, using self-organizing procedures to discover their neighbors and acquire their location. In addition, because of the dynamic nature of the network and the energy constraints of the sensors, fault tolerance and self-healing are required for an operation that will be left unattended for an extended time to be feasible. Sensors may get disconnected from the network because of their battery depletion, and new sensors may be deployed to maintain the connectivity and the environmental coverage of the network. Therefore, self-organization procedures (e.g., self-management, self-healing) must be provided so that the network can work in a robust and autonomous manner. For instance, [21]have presented a series of protocols for establishing and maintaining connectivity in WSNs: SMACS, eavesdrop and register (EAR), and sequential assignment routing (SAR) protocols. Gupta, Das, and Gu (2003) also addressed the self-organization issue and developed protocols for sensor networks for efficient query execution. Moreover, most of the data gathering approaches presented later in the chapter include methods for handling sensor node failures and insertions.

As mentioned earlier, one of the key features of self-organization is localization. Location awareness of the sensors is crucial for the operation of a large-scale network(e.g., by improving routing;[22]. In many cases, sensors are assumed to have knowledge of their location based on global positioning system (GPS) coordinates. In reality, however, this solution may be not applicable to all sensors because of cost, antenna size, and power consumption constraints; therefore, several localization techniques have been proposed. Most of these consider the case where a few sensor nodes (known as beacon or anchor nodes) of the network have a priori knowledge of their positions, either from a GPS device or by manual configuration, whereas the rest of the network sensors acquire their location by calculating their distance from the anchor nodes based on the packets received and the signal strength. Depending on the number of beacons used, as well as the algorithms deployed for

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 3521 - 3539

Received 15 December 2020; Accepted 05 January 2021.

the calculation of the distance and the angle/direction, a different grade of location accuracy can be achieved.

Experimental Sensor Network Deployment

Sensor networks can greatly improve environment monitoring, such as target detection and classification, precision agriculture, habitat monitoring, or patient monitoring. One of the experimental wireless sensor network pioneers is the Great Duck Island project (http://www.greatduckisland.net). In August 2002, researchers from UC-Berkeley/Intel Research Laboratory deployed a mote-based tiered sensor network on Great Duck Island, Maine, to monitor the behavior of storm petrel. Furthermore, in December 2004, the OSU DARPA-NEST team completed the first demonstrations and experiments of ExScal (www.cast.cse. ohio-state.edu/exscal/). The purpose of the development of ExScal is for the detection and classification of multiple intruders types over an extended perimeter. Other habitat-monitoring examples include the PODS Project (http://www.pods.hawaii.edu/), which is used to remotely monitor the rare species of plants in Hawaii for long-term study, and ZebraNet, which is used to study zebra behaviors such as long-range migration, interspecies interactions, and nocturnal behavior. The Sensor Web (http://sensorwebs.jpl.nasa.gov/) measures light levels, air temperature, humidity, soil temperature, and soil moisture, and the collected data are used to study the effects of microclimate on plant growth. Recently, Intel also deployed a sensor network with sixty-five nodes at a vineyard in British Columbia, Canada. Wine grapes are highly sensitive to temperature; therefore, real-time temperature; data from the motes (sensor nodes developed by UC-Berkeley) can identify which vines are most likely to need frost-control measures, whereas the cumulative temperature data can help the grower choose the best moment to pick the grapes. Finally, with regard to patient monitoring, several experimental sensor networks have been deployed.

The aforementioned experimental sensor networks can provide empirical data sets for further study and optimization of the data sensor networks.

Data Gathering Strategies

The simplest possible strategy to send data from the sensor nodes to a base station is called raw data gathering and consists of direct transmission of the information from all sensors to the sink. If the base station is far away from the sensors, a huge amount of transmit power is required, which would lead to a quick depletion of the energy recourses of every sensor and consequently reduce the network's lifetime. Although the raw data gathering approach could in principle provide a nearly optimal solution in cases where the sink is close to the nodes, in most cases, even for moderate-sized sensor networks, this approach is considered energy inefficient. To prolong the network's lifetime, data reduction is necessary; this is achieved mainly through in-network processing. Because the observed objects in the physical world are usually highly correlated, sensors that are deployed close to one another are expected to collect similar information about their environment, and thus data aggregation methods can be utilized to improve the overall data gathering operation.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 3521 - 3539

Received 15 December 2020; Accepted 05 January 2021.

Data gathering approaches can be classified into many diverse categories according to different features and design dimensions of the wireless sensor networks and associated protocols. More specifically, given the network's structure and organization, data gathering approaches can be divided into two main categories: hierarchical and nonhierarchical protocols. The hierarchical approaches are divided into the cluster-based, chain-based, and aggregation tree constructive protocols. In the case of WSNs, nodes are organized into groups, called clusters, and a node is elected to act as the group leader (called the cluster head). Clusters can also be used to form different layers of a hierarchy. The member nodes of each cluster send their data to the cluster head, which is responsible for sending the collected data either to a higher layer cluster head or directly to the sink. Chain-based protocols construct a chain connecting all nodes, thus reducing the total distance of data transmission. Nodes send their data to their neighbor node in the chain, and each node is responsible for forwarding its neighbor's data, possibly along with its own data. At each round, only one sensor transmits the total data packet to the sink. The hierarchical approaches include protocols that construct trees rooted at the sink and spanning the whole network. In these schemes, each node sends its data to its parent in the tree, and the parent node fuses it with its own readings and passes it to its parent one layer above until the aggregated packets reach the final destination node. On the other hand, the nonhierarchical approaches disseminate the data throughout the whole network in a flat manner, without involving any physical or logical hierarchical structure (e.g., through flooding).

One of the basic and critical operational processes in sensor networks, that is closely related to the data gathering is the routing process. Therefore, a large number of strategies that have been proposed in the literature perform the gathering of data generated in the network along with the corresponding routing decisions made, in order to optimize the overall process. On the other hand, there is a class of protocols for data gathering that aims to be routing independent. The objective of this class of protocols is to provide a more generalized and flexible data aggregation and gathering framework that achieves energy efficiency in a way that is independent of and complementary to the routing protocol. An additional dimension encountered in WSNs—different from most of other wireless and personal networks—refers to the nature and the different kinds of data that are transmitted. Although a large number of approaches send the collected data in their genuine form, it is possible that alternative approaches can be used to send encoded data to reduce the amount of data transmitted in every gathering round. This approach also constitutes a form of aggregation or, more precisely, fusion, and the corresponding techniques are known as distributed data compression techniques. They mainly reduce the amount of data transmitted by exploiting the spatial correlations that exist among different data packets.

A separate class of data aggregation methods can be created by several approaches that take advantage of the spatial correlations between the sensors' readings, in a rather different way than the ones mentioned above, by selecting only a subset of sensors to perform data gathering. More precisely, these approaches do not perform actual data aggregation but instead at each gathering round select subsets of nodes to report their readings. Neighborhood nodes are expected to collect analogous information; therefore, only one is active at each

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 3521 - 3539

Received 15 December 2020; Accepted 05 January 2021.

round and sends its gathered data to the sink. The other nodes move into idle or sleep mode to save their energy resources. These methods are comparable to the MAC approaches in which sensor nodes turn off their radios to achieve energy efficiency.

Furthermore, it should be noted that in most common scenarios in sensor networks, individual sensors are deployed in an area and are usually immobile, forming a wireless fixed network. However, there are some cases (e.g., for tracking applications) where the sensors can move either by outside force or by their corresponding mobility component. Therefore, for these cases, mobility presents an additional challenge that may affect the operation and effectiveness of the data monitoring and gathering process.

In wireless sensor networks, the sensors either continuously sense their environment and send their data to the sink in a periodic manner or gather data in an event-driven way, in which sensors report their readings only if an event has occurred (e.g., a value threshold has been exceeded). However, in some cases, users or applications may request data on demand by posing different types of queries to the sink. Therefore, the methodology and frequency of the data collection process also pose different and interesting design challenges in WSNs.

The aforementioned categorizations and classes of data monitoring and gathering approaches do not constitute mutually exclusive groups, and, as a result, one methodology may belong to more than one class. Next, we present a more detailed description of each one of these classes, by providing representative protocols and describing the operation and functionality of these approaches.

Hierarchical Protocols

Cluster-Based Protocols

As mentioned before, in cluster-based approaches clusters are formed with one leader (cluster head) at each cluster. The cluster head is engaged in collecting all the data from the members of its cluster, performing some sort of data aggregation in order to reduce the data size, and forwarding it either to a higher-layer cluster head or directly to the sink. The corresponding cluster-based formation is illustrated in Figure 1.

[19]presented the LEACH protocol, which is considered one of the most representative cluster-based protocols. In the implementation of LEACH, sensors are organized into clusters with one node acting as leader at each round. The member nodes of each cluster send their data to their cluster head, which in turn performs local data fusion to compress the amount of data to be sent, and at the end of each round sends the corresponding data to the sink. The cluster head sensors, which at each round transmit to the sink, consume significantly larger amount of energy, especially when compared to the other types of nodes. The clusters are reconstructed every round and every time a new sensor is elected cluster head in a random way. In LEACH–C the cluster formation is done at the beginning of each round using a centralized algorithm initiated by the base station. Although this version of the protocol performs better than LEACH, the associated energy cost for the cluster formation is higher and knowledge of the network topology is required for the cluster formation phase.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 3521 - 3539

Received 15 December 2020; Accepted 05 January 2021.

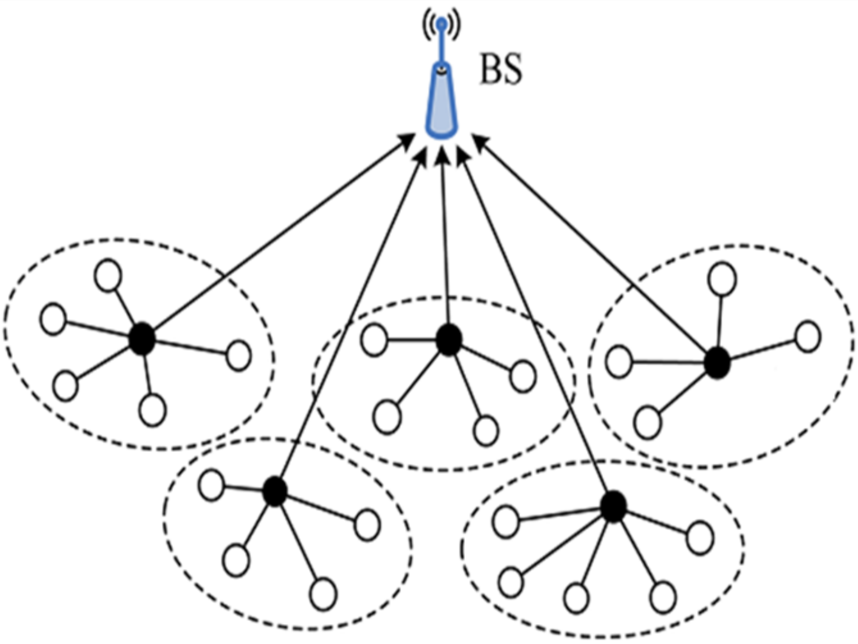

Fig. 1 Cluster-based data gathering approach

Another approach suggests an enhanced version of the LEACH algorithm, namely E-LEACH. This implementation contains four phases: a) advertisement, b) cluster setup, c) schedule creation, and d) data transmission. The first three phases are identical to LEACH algorithm, whereas in the last phase all cluster heads, after receiving the data from their cluster members, form a chain using a greedy algorithm and transmit their data along the chain.

[13]proposed an alternative cluster-based approach for data gathering. The network is partitioned into clusters, called super sensors, which make use of a greedy clustering algorithm that selects the farthest sensor node i from the sink and forms a cluster that includes node i and its (c-1) nearest neighbors, where c is a constant. The process continues until all sensors have become members of a cluster. For every super sensor, a maximum data gathering lifetime schedule is computed using a greedy clustering-based maximum lifetime data aggregation (CMLDA) heuristic. Based on each schedule computed, aggregation trees are constructed for the sensors.

Chain-Based Protocols

Another class of hierarchical protocols for data gathering is the chain-based protocols. These protocols construct a chain that connects all nodes. A representative chain-based topology is shown in Figure 2. Among the simplest chain-based protocols presented in the literature is the linear-chain scheme. The most representative examples of a linear-chain protocol is the PEGASIS protocol, in which each node communicates only with its closer neighbor and takes turns transmitting to the sink. The nodes are organized to form a chain. At each round, only one node is assigned to transmit the total data packet to the sink. Each time a different node is selected in order to increase the network’s lifetime. The drawback to the linear-chain schemes

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 3521 - 3539

Received 15 December 2020; Accepted 05 January 2021.

is the large delay and therefore enhanced chain-based schemes for data gathering have been proposed.

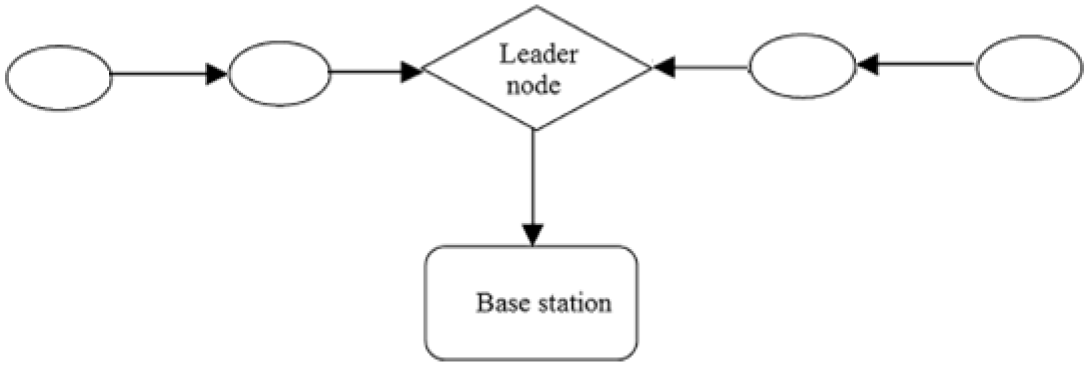

Fig. 2 Chain-based data gathering approach

A chain-based protocol was also the subject of a similar study presented by Du, Wu, and Zhou (2003), which is suitable especially in cases of sparse sensor networks. The authors divided the network into regions based on a center node. At each region, the linear-chain scheme is employed to gather the data of the sensors within the region at the center node. The center node can combine the data using an aggregation function or simply relay the separate packets to the sink. The multichain scheme proposed constructs the sub-chains through a sequence of insertions. [23]also presented a binary combining scheme using code division multiple access (CDMA), thereby avoiding radio interference. Data are combined using pairs of nodes at each level, which results in a hierarchy of$\lceil \log N \rceil$ levels, where N is the total number of sensors in the network and $\lceil \log N \rceil$represents the least integer greater than log N. For data gathering, each node at a given level transmits to its neighbor.

At the top level only one node is active and is responsible for transmitting the total packet to the base station. The nodes perform data fusion at each level, except for the end nodes. The delay cost is reduced in comparison with the simple linear-chain scheme. However, using different codes for the communication of the sensors may not be applicable, because the CDMA-capable nodes are expensive. To allow simultaneous transmissions with minimum interference, the sensors of the network are divided in G groups (where G is a random integer and G=10 is considered a near optimal choice). Within each group, the linear-chain scheme is used for G simultaneous transmissions.

Tree-Constructive Protocols

In these schemes, a rooting tree is constructed that spans the whole network. The sink initiates the process, and, at each step, new sensors join the tree until all nodes become members of the tree, either as internal nodes or leaves. The data gathering is performed along with the rooting; each node sends its data to its parent in the tree until the packets reach the destination node (i.e., the sink). Packets can also be aggregated in their way up to the root to

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 3521 - 3539
Received 15 December 2020; Accepted 05 January 2021.

conserve energy. Figure 3presents the resulting topology of a typical aggregation tree-based protocol.

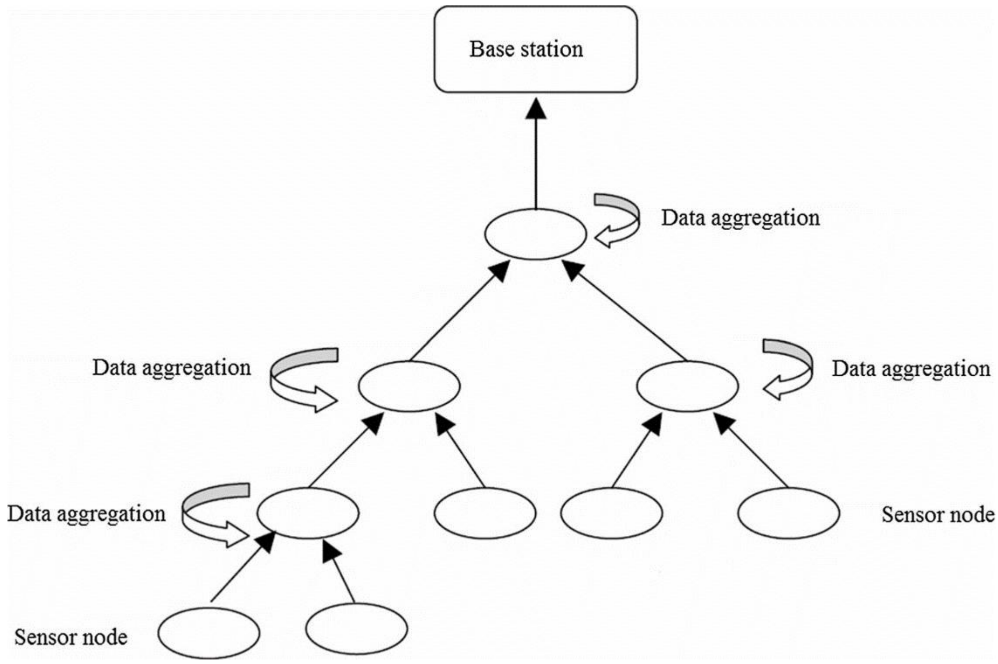

Fig. 3 Tree-constructive data gathering approach

[20]suggested the EDGEprotocol, which is basically a tree-constructing algorithm for data gathering. The tree is constructed as follows: the root initiates the tree construction by broadcasting a child request (CRQ) packet. Each node that is not a member of the tree collects a number of parent candidates, which are saved at a parent candidate (PC) table, and chooses one according to some metrics (e.g., the response time that signifies their distance). Then it sends a child reply (CPR) packet to the selected parent, and the parent responds by sending a child acceptance (CAC) packet. Specific joining and leaving procedures are provided by the protocol to handle the dynamic nature of sensor networks. If there is node failure, the tree is reconstructed.

In a similar study, [17]proposed another tree-constructive algorithm, namely the PEDAP. The PEDAP computes a minimum spanning tree over the sensor network where the costs of the edges are proportional to the transmission costs. Each sensor node belonging to the tree aggregates (or fuses) the data provided by its children with its own, and transmits one single packet to its parent until it reaches the root of the tree (i.e., the sink). Prim's algorithm is used to compute the MST, and it is initiated every k rounds to adapt to the given state of the network. Parameter k is configurable and presents a trade-off between complexity and accuracy.

Non-Hierarchical Protocols

Flooding is considered one of the most well-known nonhierarchical protocols in which the information is disseminated to the entire sensor network. The use of flooding for data gathering is considered energy inefficient because a significantly large number of redundant

Received 15 December 2020; Accepted 05 January 2021.

data packets inundate the network and sensors consume their energy to handle these packets. Therefore, this approach is rarely used for data gathering and is useful only for very specific applications. Gossiping, [24]is an enhancement of the flooding approach used to partially overcome the associated energy inefficiency problem. In gossiping, sensors do not forward the data packet to all of their neighbors but randomly select one to send the packet to. In that way, fewer copies of data packets are made and less traffic is generated in the network; as a result, less energy is depleted in every node.

Data Gathering Based on the Nature of Data

The data gathering approaches in sensor networks can be described and classified based on the nature of the data to be transmitted. Distributed data compression schemes are widely used for sending out the data collected at each gathering round. The nature of data in this case is encoded. Other techniques are based on the dissemination of metadata before sending their actual data. Distributed Data Compression Techniques One of the basic principles behind the minimization of the energy consumption in a WSN is to reduce the number of bits transmitted at each data gathering round. Another approach for achieving efficient data gathering is based on the source coding principle. There exist several distributed data compression techniques in the literature that compress the data generated at each node while exploiting the spatial correlations among them. However, any approach used for data coding and compression has to take into account the associated overhead cost for processing (i.e., coding and decoding). The processing cost should not be very high; otherwise, the energy savings from the information communication reduction would not improve the overall energy cost.

An approach that makes use of the side information presence in one node is described by Chou, where the removal of the inherent correlation in the sensors is achieved through a distributed compression algorithm. Simple lightweight encoders exist in every sensor node, and a more complex decoder exists at the gathering node (i.e., the sink). The nodes do not need to know the correlation structure in order to encode their data; they only need to have knowledge of the total number of bits that will be used for encoding. This kind of information is provided by the sink, which has global knowledge of the correlations that exist among the sensors of the network. This approach can be combined with other energy-saving techniques, such as aggregation of data, resulting in a more energy-efficient strategy for data gathering. Furthermore, [16], proposed an energy-efficient distributed coding scheme called EEADSC exploits the spatial correlation in data collected from nodes forming a cluster, based on the use of a Lagrangian cost function. This approach aims to compensate for the high energy cost incurred by the coding and decoding processing, by not allowing the decoder to directly communicate with the encoder. The EEADSC coding scheme uses trellis-coded quantization (TCQ), which results in the reduction of the bits transmitted and can be also combined with other data aggregation techniques to achieve greater energy savings.

Dissemination of Metadata

SPIN protocols belong to a family of adaptive negotiation-based information dissemination methodologies suitable for WSNs. According to this paradigm, sensors use metadata to

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 3521 - 3539

Received 15 December 2020; Accepted 05 January 2021.

describe their sensing data, whose size is small compared to data's size. Every time a sensor has data to send, it advertises it to its neighbors by sending an Advertisement message packet containing only the metadata. Then, the nodes that have received the ADV packet and are interested in the data advertised send back a request for data (REQ) message. Finally, the node sends its actual sensed data. In an enhanced energy-aware implementation of the protocol, a low-energy threshold is defined for each sensor. It reduces its participation in the whole procedure of the protocol, meaning that it will not initiate the three stage handshake if it does not have sufficient energy to complete all of the three stages. In this way, data are gathered to the BS in each round and result in an energy-aware gathering paradigm.

Inquiry-Based Data Gathering

In WSNs, usually the sensor nodes collect information about their environment (e.g., measuring the temperature for a given region) and send their readings to the sink either continuously, periodically, or whenever an event occurs that triggers such data dissemination (e.g., a value threshold has been exceeded). The majority of the protocols previously described have assumed this type of data collection process. However, in some cases on-demand data gathering may be required. This happens when a user outside the sensor network desires to collect data for a specific task by sending a declarative query to the sink. The sink proceeds in transmitting the query to all the nodes that are responsible for providing an answer to the query. The nodes send their readings back to the sink through a multi-hop route, and the intermediate nodes might perform some sort of aggregation to the data.

Temporal coherency-aware in-network aggregation, provides an improvement in terms of energy savings with some reduction in the quality of the data. A routing tree rooted at the sink is used for the propagation of the query and the collection of the results. Another unique feature of this approach is that it introduces a tolerance clause (tct) into the query, which represents the maximum change that can occur in the overall quality of data (and is defined by the user). For example, a tct of 5 percent signifies that values with changes lower than 5 percent will not be reported and calculated for the final result of the query. Consequently, energy reduction is achieved because fewer data are transmitted with the loss of the corresponding quality of data. Depending on the desired accuracy of the results, different levels of energy reduction can be provided.

Finally, the APTEENprotocol, also makes use of queries posed by users to gather the data generated in the network. This scheme categorizes the queries into three types, depending on the type of data—historical, on-time, and persistent queries—all of which are answered by the sink. The clustering algorithm used is used to partition the network into clusters, and cluster heads are charged with the aggregation of data. Furthermore, adjacent nodes that sense similar data form pairs. The APTEEN protocol achieves lower energy consumption by allowing only one sensor of every pair to send data at each round, letting the other go into sleep mode.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 3521 - 3539

Received 15 December 2020; Accepted 05 January 2021.

Selecting Subsets of Sensors for Data Gathering

Data gathering in sensor networks has been proven to be a costly operation, because sensors consume a great amount of energy when receiving and transmitting data. All the approaches presented above used different methodologies and aggregation functions to reduce data size. However, another important category of data gathering protocols exploits the data correlation provided from sensor readings that are placed close to each other, but from a different perspective. At each round, not all sensors need to send their data to the sink, and thus only a subset of nodes should be selected to transmit.

Following this paradigm, [25]investigated the problem of selecting a connected correlationdominating set that can be used for data gathering. The approach first finds the correlation existing among the sensors and afterward enforces an algorithm for finding the connected correlated subset of nodes. Similar to this approach, [20]proposed the use of data reporters for transmitting data to the sink at each round. The idea is to allow only a set of *k* sensors, named data reporters, to send data to the sink while the remaining cache their readings and send them during the following rounds, thus saving energy. Every node takes turns being selected as a data reporter. The number *k* of data reporters is selected to be sufficient for a desired sensing coverage defined by the users or the applications. The coverage is inversely proportional to the energy savings. For the selection of the *k* data reporters in each round, three schemes were developed: the non-fixed randomized selection (NRS) as well as the nonfixed and fixed disjoint randomized selection (N-DRS and F-DRS, respectively). In the first case, the selected *k* sensors in one round may not be different in the next round, whereas in the latter case, the set of *k* data reporters in a given round is completely different from the set selected in the next rounds.

Data Gathering in Mobile Environments

All of the approaches for data gathering in WSNs presented earlier in this chapter have considered cases in which the sensor nodes are mainly static and have little or no ability to move. Indeed, this is the most common scenarios in sensor networks, where sensors are deployed in an area and are usually immobile, thereby forming a wireless fixed network.

The formation can be performed by first electing the cluster heads and having these nodes broadcast advertisement messages to all the nodes, including their position, speed, and direction. Sensors in turn calculate the distance between themselves and all cluster heads and, taking into consideration their relative direction, decide which cluster to join. Because of the high power consumption and to reduce the probability that the energy of a few sensors get depleted quickly and unfairly, various algorithms for cluster head selection that will minimize the total amount of energy consumed at each round have been proposed in the literature. Experiments have demonstrated that this approach provides a power-efficient data gathering strategy for mobile sensor networks. The members of a cluster perform their tasks (e.g., sensing, data dissemination, etc.) at specific intervals and during their idle time can be put into sleep mode to save energy. Each relay node broadcasts a message Relay node ID (RID)

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 3521 - 3539

Received 15 December 2020; Accepted 05 January 2021.

with lifetime to all of the sensor nodes in a cluster. Slightly before the lifetime expires, they wake up and perform their tasks until another RID with larger lifetime broadcast occurs.

Relative Performance Comparison

In the following section, we present a relative qualitative comparison of the different data gathering approaches and highlight their respective merits, drawbacks, and performance. Specifically, we summarize the functionality and outline the basic operational characteristics of each category, and then we present a table that provides a comprehensive comparison of the various approaches against different performance metrics of interest that have been described in this chapter. The various data gathering and monitoring approaches that have been presented take into account different design objectives and principles, and therefore aim to optimize different performance parameters and metrics (e.g., energy consumption, delay/ latency, accuracy/loss of data, etc.).

More specifically, with respect to the hierarchical and nonhierarchical strategies, the protocols that belong to the first category present in principle better energy efficiency. As a result, in all the methods that belong to the hierarchical approaches (i.e., cluster-based, chain-based, and tree-constructive protocols), the network lifetime is increased at the cost of the delay. Furthermore, associated with the hierarchical approaches is an initialization phase that is required for the definition of the different layers of hierarchy, as well as maintenance costs required for the management and reconstruction of the hierarchical approach during the data gathering operation. It has been also demonstrated that the energy cost introduced by data aggregation is negligible compared to the corresponding communication and data transmission cost. Depending on the data aggregation function used at each sensor node (e.g., average, sum, discard of duplicate packets), the final data delivered to the collection center may be different from the original data, resulting in some level of loss in the data accuracy.

On the other hand, the nonhierarchical approaches (e.g., flooding, gossiping, etc.) do not need to go through any initialization phase and in general present lower implementation complexity and maintenance costs. All of the sensor nodes that are involved in the data gathering operation send their data to all of their neighbors without performing any in-network processing (e.g., data aggregation). As a result, all of the sensors consume larger amount of energy and the corresponding network lifetime is decreased. With regard to consideration of the delay constraints, the nonhierarchical protocols deliver data in a timelier manner during their first periods of operation or when they operate under low traffic load. However, as the traffic load increases and the data packets that travel through the network increase, severe delays are observed from collisions and bottlenecks; this also has a significant negative impact on the achievable network throughput.

With respect to the probabilistic and routing independent approaches, in addition to their distributed nature and demonstrated energy efficiency, one of their key principles is that they can be combined with any other energy-aware routing protocol to attain even higher energy gains. Furthermore, because of their distributed, probabilistic behavior, they can decide independently whether or not to perform in-network processing to reduce the amount of data transmitted, thus succeeding in satisfying the delay constraints posed by the application.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 1, 2021, Pages. 3521 - 3539

Received 15 December 2020; Accepted 05 January 2021.

The data gathering approaches that are based on the nature of data mainly aim to accomplish the data gathering operation in an energy-efficient manner. More specifically, the protocols that perform data compression in the presence of correlated data transmit fewer amounts of data with the trade-off of the occurrence of some information loss (provided that the compression is not lossless), while the use of encoders and decoders increases the processing cost and latency. If metadata is transmitted, there is no additional processing cost and the total amount of data packets traversing the network is reduced. With reference to the inquiry-based data gathering protocols, the collection center poses different queries to the network that are addressed only by a subset of nodes, thus reducing the total amount of information collected at the center. The accuracy of the collected data depends on whether or not the nodes perform some kind of aggregation as well as the type of the aggregation function utilized. Moreover, in certain protocols, sensor nodes do not transmit their reading if there is a small variation from the previous transmission, thus resulting in greater information loss in some cases.

Finally, the last class of protocols considered here does not perform in-network processing but explores the correlation of the sensor readings by not allowing some sensors to transmit in specified communication rounds. The main characteristic of these approaches is that the energy consumption can be controlled (i.e., reduced) by adjusting the subset of sensors to transmit at each cycle, at the cost of reducing the corresponding achievable network sensing coverage. Table 1presents a qualitative relative comparison of the aforementioned methods against a common set of parameters.

CONCLUSION

In this paper, we focused on the problem of energy-efficient data monitoring and gathering in sensor networks. The sensor networks should dynamically adapt to the system and topology changes, and at the same time they need to balance the trade-offs among various performance metrics. Therefore, we first identified the major performance metrics of interest for the data monitoring and gathering process, and we discussed the related traffic and data correlation models. We then identified and reviewed various issues associated with the sensor node deployment patterns, which are closely related to the effective data monitoring. Furthermore, we described several data gathering approaches by classifying them into diverse categories according to the different features and design dimensions of the WSNs, and we presented some indicative associated protocols. Finally, we provided a relative comparison of all of the different approaches—outlining their major merits and drawbacks, highlighting their differences, and evaluating their relative behavior based on a common set of parameters.

REFERENCES

[1] G. Kadiravan *et al.*, “Metaheuristic clustering protocol for healthcare data collection in mobile wireless multimedia sensor networks,” *Comput. Mater. Contin.*, vol. 66, no. 3, pp. 3215–3231, 2021.

[2] S. R. Hassan, I. Ahmad, S. Ahmad, A. Alfaify, and M. Shafiq, “Remote pain

Received 15 December 2020; Accepted 05 January 2021.

monitoring using fog computing for e-healthcare: An efficient architecture," *Sensors (Switzerland)*, vol. 20, no. 22, pp. 1–21, 2020.

[3] B. Goyal, A. Dogra, S. Agrawal, B. S. Sohi, and A. Sharma, "Image denoising review: From classical to state-of-the-art approaches," *Inf. FUSION*, vol. 55, pp. 220–244, Mar. 2020.

[4] G. Gardašević, K. Katzis, D. Bajić, and L. Berbakov, "Emerging wireless sensor networks and internet of things technologies—foundations of smart healthcare," *Sensors (Switzerland)*, vol. 20, no. 13, pp. 1–30, 2020.

[5] Z. Shahbazi and Y.-C. Byun, "Towards a secure thermal-energy aware routing protocol in wireless body area network based on blockchain technology," *Sensors (Switzerland)*, vol. 20, no. 12, pp. 1–26, 2020.

[6] J. S. Chohan *et al.*, "Mechanical Strength Enhancement of 3D Printed Acrylonitrile Butadiene Styrene Polymer Components Using Neural Network Optimization Algorithm," *Polymers (Basel).*, vol. 12, no. 10, p. 2250, Sep. 2020.

[7] P. K. Sahoo, S. R. Pattanaik, and S.-L. Wu, "A novel synchronous MAC protocol for wireless sensor networks with performance analysis," *Sensors (Switzerland)*, vol. 19, no. 24, 2019.

[8] M. Faheem *et al.*, "A multiobjective, lion mating optimization inspired routing protocol for wireless body area sensor network based healthcare applications," *Sensors (Switzerland)*, vol. 19, no. 23, 2019.

[9] Y. Qu, G. Zheng, H. Ma, X. Wang, B. Ji, and H. Wu, "A survey of routing protocols in WBAN for healthcare applications," *Sensors (Switzerland)*, vol. 19, no. 7, 2019.

[10] A. R. Bhangwar *et al.*, "WETRP: Weight Based Energy Temperature Aware Routing Protocol for Wireless Body Sensor Networks," *IEEE Access*, vol. 7, pp. 87987–87995, 2019.

[11] P. K. Sahoo, S. R. Pattanaik, and S.-L. Wu, "A reliable data transmission model for IEEE 802.15.4e enabled wireless sensor network under wifi interference," *Sensors (Switzerland)*, vol. 17, no. 6, 2017.

[12] D. Sathya and P. G. Kumar, "Secured remote health monitoring system," *Healthc. Technol. Lett.*, vol. 4, no. 6, pp. 1–5, 2017.

[13] H. Fouad and H. Farouk, "Heart rate sensor node analysis for designing internet of things telemedicine embedded system," *Cogent Eng.*, vol. 4, no. 1, 2017.

[14] U. Singh, R. Salgotra, and M. Rattan, "A Novel Binary Spider Monkey Optimization Algorithm for Thinning of Concentric Circular Antenna Arrays," *IETE J. Res.*, vol. 62, no. 6, pp. 736–744, 2016.

[15] M. K. Gupta *et al.*, "Parametric optimization and process capability analysis for machining of nickel-based superalloy," *Int. J. Adv. Manuf. Technol.*, vol. 102, no. 9–12, pp. 3995–4009, Jun. 2019.

[16] C. Li, X. Yuan, L. Yang, and Y. Song, "A hybrid lifetime extended directional approach for WBANs," *Sensors (Switzerland)*, vol. 15, no. 11, pp. 28005–28030, 2015.

[17] E. Kartsakli, A. Antonopoulos, L. Alonso, and C. Verikoukis, "A cloud-assisted random linear network coding medium access control protocol for healthcare applications," *Sensors (Switzerland)*, vol. 14, no. 3, pp. 4806–4830, 2014.

[18] J. I. Bangash, A. H. Abdullah, M. A. Razzaque, and A. W. Khan, “Reliability aware routing for intra-wireless body sensor networks,” *Int. J. Distrib. Sens. Networks*, vol. 2014, 2014.

[19] X. Y. Wang and A. Wong, “Multi-Parametric Clustering for Sensor Node Coordination in Cognitive Wireless Sensor Networks,” *PLoS One*, vol. 8, no. 2, 2013.

[20] C.-G. He, S.-D. Bao, and Y. Li, “A novel tri-factor mutual authentication with biometrics for wireless body sensor networks in healthcare applications,” *Int. J. Smart Sens. Intell. Syst.*, vol. 6, no. 3, pp. 910–931, 2013.

[21] H. Mousannif and I. Khalil, “Cooperative Data Management in Wireless Sensor Networks,” *Int. J. Comput. Intell. Syst.*, vol. 5, no. 3, pp. 403–412, 2012.

[22] P. Kumar, S.-G. Lee, and H.-J. Lee, “E-SAP: Efficient-strong authentication protocol for healthcare applications using wireless medical sensor networks,” *Sensors*, vol. 12, no. 2, pp. 1625–1647, 2012.

[23] T. H. B. Singh, J. S. Chohan, and R. Kumar, “Performance analysis of vapour finishing apparatus for surface enhancement of FDM parts,” *Mater. Today Proc.*, vol. 26, pp. 3497–3502, 2020.

[24] S. Tennina *et al.*, “WSN4QoL: A WSN-oriented healthcare system architecture,” *Int. J. Distrib. Sens. Networks*, vol. 2014, 2014.

[25] M. K. Khan and S. Kumari, “An improved user authentication protocol for healthcare services via wireless medical sensor networks,” *Int. J. Distrib. Sens. Networks*, vol. 2014, 2014.