A QoS-Aware Data Collection Protocol for LLNs in Fog-Enabled Internet of Things

A. S. M. Sanwar Hosen , Saurabh Singh , Pradip Kumar Sharma, Md. Sazzadur Rahman,
In-Ho Ra, Gi Hwan Cho, and Deepak Puthal , *Senior Member, IEEE*

Abstract—Improving quality of service (QoS) of low power and lossy networks (LLNs) in Internet of things (IoT) is a major challenge. Cluster-based routing technique is an effective approach to achieve this goal. This paper proposes a QoS-aware clustering-based routing (QACR) mechanism for LLNs in Fog-enabled IoT which provides a clustering, a cluster head (CH) election, and a routing path selection technique. The clustering adopts the community detection algorithm that partitions the network into clusters with available nodes' connectivity. The CH election and relay node selection both are weighted by the rank of the nodes which take node's energy, received signal strength, link quality, and number of cluster members into consideration as the ranking metrics. The number of CHs in a cluster is adaptive and varied according to a cluster state to balance the energy consumption of nodes. Besides, the protocol uses the CH role handover technique during CH election that decreases the control messages for the periodic election and cluster formation in detail. An evaluation of the QACR has performed through simulations for various scenarios. The obtained results show that the QACR improves the QoS in terms of packet delivery ratio, latency, and network lifetime compared to the existing protocols.

Index Terms—Internet of Things, fog computing, low power and lossy network, clustering, routing, quality of service.

I. Introduction

INTERNET of Things (IoT) has become a mature technology that go beyond simple, fine-grained safety or event-monitoring platforms. IoT generally refers to the interconnection of things, a form of low power and lossy network (LLN). LLNs are deployed in different IoT applications [1]. A LLN is a collection of resource-constrained sensor devices deployed in an area of interest for sensing and gathering data. The collected data are sent to one or more control station called base station (BS) or gateway (GW). In a Fog-enabled IoT, the collected date are stored in a Fog server for pre-processing before sending it to cloud, so that computation, storage, and networking services can be performed locally [2], [3]. The data of LLNs are distributed for several services in IoT applications.

A Fog-enabled IoT consists of three-layer architecture, that is, the LLNs of sensor nodes that is deployed on the network edge, the Fog-enabled BSs/GWs that aggregate data from LLNs and perform various operations, and the IoT middleware (MW) plays a role as a back-end cloud to provide services. The resources of the sensor devices used in a LLN/wireless sensor network (WSN) are constrained with regard to battery power, processing, storage capacity, and communication bandwidth. To address this issue, different sensor node deployment optimization [4], [5] and data collection protocols [6] have been proposed. Some of the protocols are application specific [7]–[11], while others are for general application [12]. Routing protocols based on the clustering are considered an energy-efficient technique in terms of data aggregation and sending to the BS [13]. In clustering, the deployed nodes are partitioned into groups called clusters and cluster heads (CHs) are elected in the clusters. A CH aggregates data from its cluster members (CMs) and eliminates correlated data to reduce the amount of data sent to the BS.

The clustering techniques can be classified into three main categories: static, dynamic, and hybrid under centralized or distributed process [18]–[40]. Static clustering is predefined, and it divides the network into levels, grids, regions, sectors, and so on [30], [37]–[40]. Dynamic clustering is usually self-configured (can be AI-assisted) and random according to the network state [18]–[29], [31]–[33]. On the other hand, hybrid clustering adopts either static or dynamic clustering according to what is needed [34]–[36]. Sensor nodes are deployed uniformly or non-uniformly in LLNs/WSNs. In most clustering techniques, a node needs to detect which cluster it belongs to. For this, either received signal strength (RSS) or global positioning system (GPS) is usually used. RSS may not serve the purpose when the environment is considered very noisy. On the other hand, each node needs to be equipped with GPS in order for its geographical location to be obtained, which might not be applicable in some environments such as inside a building or cave.

A. S. M. S. Hosen and G. H. Cho are with the Division of Computer Science and Engineering, Chonbuk National University, Jeonju 54896, South Korea (e-mail: sanwar@jbnu.ac.kr; ghcho@jbnu.ac.kr).

S. Singh and I.-H. Ra are with the School of Computer, Information and Communication Engineering, Kunsan National University, Gunsan 54150, South Korea (e-mail: singh1989@jbnu.ac.kr; ihra@kunsan.ac.kr).

P. K. Sharma is with the Department of Multimedia Engineering, Dongguk University, Seoul 04620, South Korea (e-mail: pradip.academic@gmail.com).

M. S. Rahman is with the Institute of Information Technology, Jahangirnagar University, Dhaka 1342, Bangladesh (e-mail: sazzad@juniv.edu).

D. Puthal is with the School of Computing, Newcastle University, Newcastle upon Tyne, U.K. (e-mail: dputhal88@gmail.com).

Digital Object Identifier 10.1109/TNSM.2019.2946428

Clustering without knowing the nodes' connectivity might not be efficient in LLNs due to lack of knowledge of location and link quality of the nodes. These problems may become severe when the environment is uneven and noisy. For example, if nodes are non-uniformly distributed over a sensor field, the number of CMs in the clusters varies significantly. This results in an imbalance in the energy consumption of the nodes. Meanwhile, the lack of knowledge about the link quality in CH election may increase the packet drop probability during communication. This may cause an unreliable communication and cause nodes to dissipate some energy in the data retransmission process. These greatly impact the QoS of the network, particularly with regard to the packet delivery ratio (PDR) and the network lifetime (NL).

To address the aforementioned problems, a QoS-aware clustering-based routing (QACR) protocol for LLNs in Fog-enabled IoT is proposed. The protocol adopts the community detection technique in clustering, by which the available nodes' connectivity information is utilized. The CH election and relay node selection are based on the rank. The rank of the nodes is weighted by the residual energy, RSS, link quality and the number of CMs. Data routing is hierarchical from a CM to the BS based on the constructed routing path. This reduces the packet loss, latency, and ensures a balanced energy consumption of the nodes. The major issues are taken into consideration in clustering and routing to improve the QoS of the network.

The remaining of the paper is organized as follows. Section II discusses the data collection protocols using clustering. Section III presents the system model. Section IV describes the details of the proposed data collection technique QACR. The performance of QACR is evaluated in Section V. The concluding remarks of this work are given in Section VI.

II. Cluster-Based Routing Protocols

The clustering of sensor nodes is usually adopted in large-scale networks. Cluster-based networks provide more reliability, better coverage, greater fault tolerance, and better task allocation and energy-efficiency [13]–[17]. Several cluster-based routing protocols for LLNs/WSNs have been well-studied and proposed in the last decade in attempts to resolve the "energy-hole" problem [12]. The low-energy adaptive clustering hierarchy (LEACH) [18] is in this category. The idea of LEACH is the clustering of nodes on independently elected CHs with a probability and rotating of the CH role to balance the energy consumption. However, it requires a large control-message overhead resulting in additional energy dissipation of the nodes. Inspired by LEACH, many protocols have been proposed in which different weight functions are considered in the CHs and route-selection processes to improve the performance over the protocol. The centralized LEACH (LEACH-C) [19], multi-level clustering EEMLC [20], EADEEG [21], RDBC [22], and EDDEEC [23] are examples of routing protocols in the LEACH family [24]. In EADEEG, the CH election is based on the ratio of the average residual energy of neighbors to the residual energy of a node itself. This balances the energy consumption of nodes and prolongs the NL over the LEACH. The drawback of this protocol is that some of the nodes may be isolated from the rest of the nodes.

The authors in [25] proposed an energy-aware distributed unequal clustering (EADUC) protocol. It addresses the node-isolation and NL issue. The adopted CH election with a probability at time T is based on the ratio of the residual energy to the average residual energy of neighbors of a node. A CH is elected in each competition range R_c and broadcasts a head message within a range $R_r = 2R_c$. The R_c is varied according to the distance from the BS. If an exception happened, a node which has not received the head message from a tentative CH is elected as a CH independently. A cluster is formed with the nearby nodes defined by the distance from a CH. A relay CH is selected based on the shortest distance from the BS. Unlike previously discussed protocols, EADUC achieves an enhanced NL. The authors in [26] studied a k-connected overlapping technique in cluster formation to solve the coverage and NL issue. The CH election is based on the available energy status of the nodes. It achieves a prolonged NL compared to the other protocols. A NL maximization protocol has been proposed in [27]. When electing a CH, this protocol considers the available energy of nodes and the required energy consumption of the route towards the BS. It distributes the load of data routing to the nodes. The authors in [28] proposed a link-aware clustering protocol for event driven WSNs. In CH election, this protocol employs the predicted transmission count as a clustering metric. It ensures a reliable and energy-efficient route in data communication.

An energy-efficient routing protocol for non-uniform node distribution has been proposed in [29]. The network is partitioned into equal clusters defined by a competition range R_c. The CH election and clustering policy of this protocol are similar to that of EADUC. The criterion of a relay CH selection is the higher relay value defined by the residual energy and the number of CMs in each cluster. It achieved a more prolonged NL than either the LEACH or EADUC protocol. A decentralized cluster-based routing protocol called DHCRA has been proposed in [30]. The main approach of this scheme is that CHs are elected at different levels along with the construction of the routing trees. The CH election is weighted by the residual energy and distance from the BS. The policy is intended to reduce the control-message overhead with regard to the route construction. But the CH election process can be disrupted due to the election policy, which may result in an unreliable communication at a certain round over the time. The authors in [34] proposed an adaptive and distributed clustering method called DARC. The main idea of DARC is to distribute the routing tasks among relay CHs by adjusting the routing mode of CHs. Although the periodic CH election is able to distribute the data collection and routing tasks among nodes, the energy consumption of nodes is imbalanced due to the random distribution of nodes. This method adopts a relay CH selection on the basis of two relay modes, CH with low energy and CH with high energy to address the problem. During the data communication, a CH with low energy selects one of the CHs with high energy as a relay node towards the BS. A CH with high energy

transmits data to the BS directly. This process distributes the routing tasks among the CHs with high energy and balances the energy consumption of the nodes. It results in a prolonged NL compared to other protocols.

Despite the conventional CH election techniques, a hybrid unequal clustering (HUCL) protocol has been proposed to reduce the control-message overhead, so that the NL is improved [35]. It suggests a CH handover and piggybacking technique. Once a node is elected as a CH, the node hands over the CH role to other in the next round. The weight value of the CMs are piggybacked along with the local data, so that the number of control-messages broadcast regarding the CH elections is reduced. It enhances the steady state, as well as the NL. The authors in [36] proposed a modification of EADUC called IEADUC. It used the CH role handover approach of HUCL and the CH election method of EADUC. Besides, it defines a modified relay function which has been used in EADUC. The adopted relay CH selection is based on the relay value, where the residual energy of a CH, number of CMs, and energy consumption in data processing to a next hop towards the BS are utilized. The IEADUC protocol achieves more enhanced NL than EADUC and HUCL.

A clustering hierarchy protocol (CHP) based on particle swarm optimization (PSO) algorithm has been proposed in [37] to improve the NL. Three different types of nodes CHs, relay nodes (RNs) and common nodes (CNs) are defined in the network. The CHs and RNs are selected by the BS using the PSO algorithm utilizing the fitness functions that include nodes' residual energy and location from the BS. A node with higher residual energy and near the BS is preferred to be selected as a CH. The clusters are formed based on the RSS on received advertisement messages from the selected CHs. A CH selects an RN towards the BS. The protocol improves the NL than that of other protocols. An energy centric cluster-based routing (ECCR) protocol has been proposed in [38]. It divides the network into a number of static grids called clusters. It adopts the handover technique inspired by the HUCL protocol, but the weight functions of CH and relay node selection differ. In CH election, the weight function includes the residual energy and the average distance among CMs. The adopted route selection includes the factors that have the major influence in energy consumption of the relay nodes. The ECCR protocol improved the NL more than the HUCL and IEADUC protocols. However, the existing protocols emphasized to prolong the NL only, where the PDR and latency of QoS have not taken into consideration.

III. SYSTEM MODEL

A. Fog-Enabled IoT Architecture

The architecture of the IoT system assumed here is three-layer consisting of LLN sensor nodes, IoT-GW, and IoT-MW as shown in Fig. 1. The LLN, where the IoT-devices are deployed and collect data locally. The IoT-GW contains the BSs/GWs that play the role of a Fog layer. This layer belongs to the nearby LLN to process data aggregated by the network. The IoT-MW plays the role of the back-end cloud with the
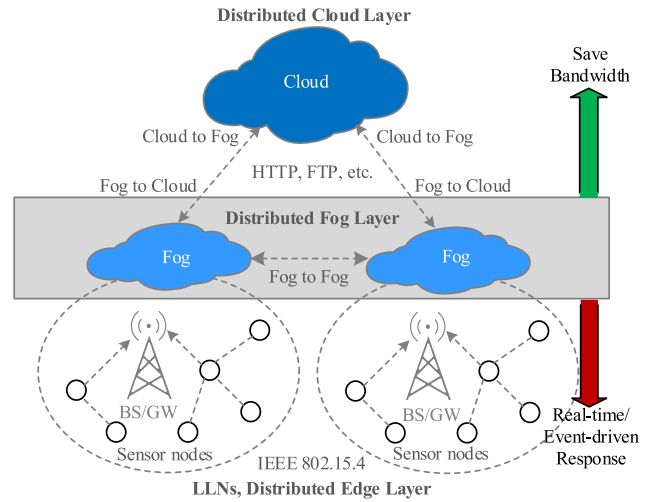

Fig. 1. An example architecture of a Fog-enabled IoT.

duty of virtualization, data storage and service providing. The components of the architecture are described subsequently.

1) LLN: It is comprised of resource-constrained sensor nodes that are connected to each other wirelessly. The specification of the devices is standardized by the IEEE 802.15.4. The sensor nodes are deployed in a LLN can be analog or digital, and able to sense data such as temperature, humidity, light, noise, and so on. The data are transmitted to the BS/GW within a range for further processing.

2) IoT-GW: The IoT-GW is positioned as a bridge between the LLN and the MW. It handles various functionalities from sensor node deployment to the data forwarding, including (de-)registration of sensor nodes, network maintenance, data collection protocols and data communication to the MW. Necessarily, it acts as a communication hub of the two other layers. Depending on the applications and the IEEE standard offered, it enables connectivity to the LLN. On the other hand, connectivity to the MW is provided by the Ethernet or ISM band (i.e., Wi-Fi, Bluetooth, ZigBee, etc.) interface of a running operating system.

A Fog-enabled GW is used to (re-)register sensor nodes and collects data from the LLN. Sensor nodes are registered to the GW by sending a registration-request message providing necessary information such as ID in IPv6. The GW stores the information in its database and sends the message to the MW. Furthermore, for a service request provided by an appropriate communication protocol (i.e., HTTP, FTP, etc.) from the MW, it manages the (de-)registration and (de-)activate of a node in the network. For time-critical applications, it processes data (i.e., optimization, averaging, pattern recognition, etc.) locally to provide QoS. Additionally, it includes various security measures and reliability in terms of connectivity between the MW and the GW. This ensures the local logging and sending data to the MW to perform uninterruptedly.

3) IoT-MW: The IoT-MW is a back-end cloud where services are requested and provided. Along with the security and privacy features, the services include data processing and availability, and virtualization of data. The implementation of

MW is depended on the application independent software, its openness features and design selection criteria defined by the IoT architecture. The functional components can be pointed as follows:

1) Handling the service requests from the applications through validation of the virtual entities that are interested of the applications.
2) Creating and managing the digital representation of the sensor nodes.
3) Creating and managing communities of sensor nodes, which offer advanced services to the users.
4) Managing and processing data (i.e., optimization, averaging, filtering, pattern recognition, etc.) based on application requirements and service request.

B. LLN Model and Assumptions

A LLN comprises of N sensor nodes $S = \{s_1, \ldots, s_N\}$ and a BS. The nodes are deployed in an area of interest M to monitor the environment. Once the nodes and the BS are deployed, they are static. Each node has an *ID* number, and nodes are location unaware. Considering the initial energy, the nodes can be homogeneous or heterogeneous. The nodes and BS are capable of adjusting their transmission power in accordance with the distance to the desired recipient [41], [42].

C. Energy Consumption of LLN Model

The radio energy dissipation of LLN model used in this paper is referred to [37], [43]. A node consumes energy in transmitting E_{TX} and receiving E_{RX} of *l*-bits data over a radio range d according to Eqs. (1) and (2), respectively. Depending on distance d, the energy consumption of a node in terms of E_{TX} and E_{RX} are defined by the free space ε_{fs} and multipath fading ε_{mp} frequency models where the d^2 and d^4 power loss are used, respectively. A CH consumes E_{da} (nJ/bit/signal) in data aggregation and E_{com} (nJ/bit/signal) in processing the values (i.e., rank and residual energy) sent by a CM.

$$E_{TX}(l, d) = \begin{cases} l \times E_{elec} + l \times \varepsilon_{fs} \times d^2 & if\, d \leq d_0 \\ l \times E_{elec} + l \times \varepsilon_{mp} \times d^4 & otherwise \end{cases} \quad (1)$$

$$E_{RX}(l, d) = l \times E_{elec} \quad (2)$$

where, E_{elec} is the power consumption of the transceiver circuits. The signal is amplified that depends on the d and the reference $d_0 = \sqrt{(\varepsilon_{fs}/\varepsilon_{mp})} \approx 87$ m.

D. Data Aggregation Model

Data aggregation compresses an amount of data into a packet of fixed-size, regardless of the number of packets received. The data compressibility model used in this work is similar to that of in [14]. In this model, the data of CMs are aggregated at their respective CH prior to transmission to the BS. During the network operation, each CM's radio is turned off until the transmission time assigned to the node. Also, it needs to be turned on for the CH until all data is received from the CMs.

TABLE I
DESCRIPTION OF THE CONTROL MESSAGES

Message	Description
Hello_Msg₁	[*selfid*, *status(BS)*], a message of self-introduction of the BS to the nodes and to measure the $RSS(BS,s_i)$.
Hello_Msg₂	[*selfid*, *RSS(BS,s_i)*], a message of a node's information and to measure the $RSS(s_i,s_j)$.
Hello_Msg₃	[*selfid*, *NAT*], a message of the neighbor's information of a node.
Hello_Msg₄	[*clusterid*, *clusterrank*, *CMsid*], a message of cluster's information.
Node_Msg	[*selfid*, *selfenergy*, *rank*, *flag*], a message of CM's information. A *flag* is the information of a CM that can communicate with the rest of the CMs in a cluster, if so *flag* = 'FC', otherwise *flag* = 'NFC'.
Handover_Msg	[*selfid*, *clusterid*, *headid*], a message to handover a CH role to a prospective CH.
Join_Msg	[*selfid*, *clusterid*], a message to join a relay node to send data to a CH in a cluster.
Schedule_Msg	[*headid*, *clusterid*, *schedule*, *order*], a message to assign the time slots for the CMs.
Route_Msg	[*headid*, *clusterank*, *noderank*, *selfenergy*, *RSS(BS,CH_i)*], a message to collect neighbor CHs' information.

IV. QACR PROTOCOL DETAILS

The QACR operation starts with the network initialization phase. In this phase, a set of information of deployed nodes such as RSS from BS, RSS from neighbor nodes, and link quality with neighbor nodes are collected. Then the nodes are partitioned into clusters. The clustering is carried out by the BS. The number of clusters and the CMs is static. The remaining operations of the protocol are conducted in regular rounds. Every round comprises the CH election and data communication phases. CMs collect local data and transmit to their respective CH. The CH aggregates data from its CMs and transmits the data to the BS using single hop or multi hop that depends on the distance of the BS. The data communication phase should be longer than the CH election phase to minimize the control-message overhead and enhance the NL. Several types of control messages used during the protocol operation are shown in Table I.

A. Network Initialization Phase

1) Information Collection: The network initialization phase starts through the information collection. At the beginning of this phase, the BS broadcasts a *Hello_Msg₁* multiple times at a certain power level over the sensor field. The power level is adjusted in accordance with the network area. Nodes receive the messages, measure the approximate $RSS(BS, s_i)$ and compute the approximate distance $d(s_i, BS)$ based on the RSS [44]. Then each node broadcasts a *Hello_Msg₂* using a radio range R_n, where R_n is less than or equal to the maximum transmission range R_{max} of a node. The nodes reside the range, receive the message, measure the $RSS(s_j, s_i)$, compute the approximate distance $d(s_i, s_j)$, and count the number of retransmissions of the message. A node lists its neighbors' information in its neighbor table (NT). Two nodes (s_i, s_j) are said to be neighbor if a node s_i receives the message from a node s_j directly.

Once the $Hello\_Msg_2$s broadcast is completed, each node calculates the expected transmission count (ETX). The channel condition of wireless links varies among nodes according to the environment. The data delivery is likely to fail through an unreliable link that leads to packet retransmissions. *ETX* is the expected number of transmissions required by a packet to be delivered successfully [45]. It is a metric of link reliability and usually used to evaluate the level of link quality. A link quality increases with decreases the value of *ETX*. The expected bi-directional transmission count of a link between node s_i and node s_j can be defined as in Eq. (3), which has been proposed in [24].

$$ETX\left(s_i, s_j\right) = \frac{1}{p_f\left(s_i, s_j\right) \times p_r\left(s_i, s_j\right)} \tag{3}$$

where, $p_f(s_i, s_j)$ and $p_r(s_i, s_j)$ represent the forward and reverse delivery ratios from node s_i to node s_j, respectively. The forward delivery ratio is the measured probability that a data packet is received successfully by the recipient. The reverse delivery ratio is the measured probability that the acknowledgement packet is successfully received. The number of packet retransmissions of a dropped packet is restricted by a threshold TH_{RC}.

Unlike the previous work, the proposed QACR defines the link quality between two nodes as '*good*' or '*bad*' that are referred by $LQ(s_i, s_j) = 1$ and $LQ(s_i, s_j) = [0, 1)$, respectively. Each node calculates the *LQ* with the neighbor nodes and updates its NT. The *LQ* is defined by the *ETX* as in Eq. (4).

$$LQ\left(s_i, s_j\right) = \frac{1}{ETX\left(s_i, s_j\right)} \tag{4}$$

When the NTs of the nodes are completed, each node short-lists its neighbors to a subset called neighbors in an adaptive range table (NAT), where an adaptive range R_{adp} is less than or equal to R_n. Each node sends its NAT along with a $Hello\_Msg_3$ to the BS. If the BS is out of R_{max} from a node s_i, the node selects a node with the higher RSS from the BS belonging to its NT as a relay node towards the BS. The packet forwarding is continued until the packet reaches the BS. The BS constructs an adjacency matrix $A(S, E)$ of size $N \times N$ based on the received information of the nodes. Fig. 2 illustrates an example scenario of the message sending from a node to the BS. As seen in the figure, RSS of the nodes from the BS are different in accordance with their locations. A node s_4 is out of R_{max} from the BS and it selects a relay node s_3 having the higher *RSS*(dBm) from the BS among the neighbor nodes s_1, s_2 and s_3.

2) Clustering and Membership: The clustering is based on the community detection algorithm [46]. A community detection algorithm is often used to detect partitions of nodes in a network (i.e., computer network, social network, biological network, etc.) that are more densely connected internally than with the rest of the network. The BS detects the communities called *K*-clusters using the modularity maximization algorithm. The modularity maximization is an optimization method of clustering that defines the number of clusters based on a quality function *Q*. The clustering process is recursively
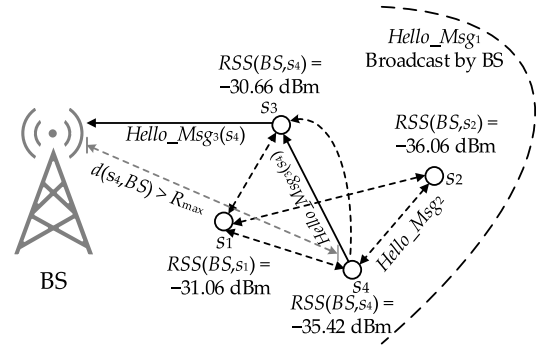

Fig. 2.   An example of nodes' information collection and sending to the BS.

partitioning the graph $G(S, E)$ into two subgraphs and repeatedly applying the same procedure to the subgraphs. The value of *K* is defined by the quality function $Q(G, C)$. When the value of $Q(G, C)$ is a constant (i.e., a value of $Q(G, C)$ is an interval of [0.5, 1]), the *K* is a reciprocal function of R_{adp}. An R_{adp} can be defined by the node density factor γ (the higher the density, the lower the value of γ). The detail of the clustering is as follows:

Let's consider a graph *G* is used as a synonym of a network; cluster as a synonym of a community; clustering as a synonym of community detection.

1) A graph $G = (S, E)$ consists of a node set *S* and an edge set *E*. The nodes contained in *S* will always be labeled as $1, 2, 3, \ldots, N$. The edges contained in *E* indicate which nodes are connected. If $e(s_i, s_j) \in E$, then we know that node s_i is connected to node s_j. So, $e(s_i, s_j)$ is the same as $e(s_j, s_i)$, they both tell us that s_i and s_j are connected.

2) The graph *G* has an adjacency matrix *A*. If the *G* has *N* nodes, *A* is an $N \times N$ matrix of 0's and 1's. If $e(s_i, s_j) \in E$ and $e(s_i, s_j) = e(s_j, s_i)$, then $A(s_j, s_i) = 1$. *A* is a symmetric matrix as follows:

$$A(S, E) = \begin{bmatrix} e(s_1, s_1) & e(s_1, s_2) & \ldots & e(s_1, s_N) \\ e(s_2, s_1) & e(s_2, s_2) & \ldots & e(s_2, s_N) \\ \ldots & \ldots & \ldots & \ldots \\ e(s_N, s_1) & e(s_N, s_2) & \ldots & e(s_N, s_N) \end{bmatrix}$$

3) A clustering of $G(S, E)$ is a partition of *S* into sets $C = c_1, \ldots, c_K$ such that $c_1 \cup \cdots \cup c_K = S, c_1 \cap \ldots \cap c_K = \{\}$ and none of $c_1, \ldots, c_K$ is empty. The sets $c_1, \ldots, c_K$ are the clusters. We write the partition as $C = \{c_1, \ldots, c_K\}$. The size of partition is $K = |C|$.

4) Given a graph $G = (C, E)$ and a partition $C = \{c_1, \ldots, c_K\}$, the edges of *G* can be partitioned into sets E_{ij} as $e(s_i, s_j) \in E_{kl}$ (if and only if $s_i \in c_k$ and $s_j \in c_l$, where $(k, l) = 1, \ldots, K$ and $k \neq l$).

5) In particular, we write $E_k^i = E_{kk}$ and $E_k^g = \cup_{l \neq k} E_{kl}$. In other words, the set E_k^i contains the internal edges of c_k, with both their ends belonging to the same cluster, while the set E_k^g contains the external edges of c_k, which have one end and the other end in $S - c_k$, the set of nodes which do not belong to c_k.

6) A community can be defined in terms of a quality function. A quality function is a function $Q(G, C)$ (i.e., it depends on the both graph *G* and *C*), the value

of which characterizes how good C is as a partition of G. Hence the best composition of G into communities is the partition $C^* = \{c_1^*, \ldots, c_K^*\}$ which maximizes Q, for instance, $C^* = argmax_C Q(G, C)$. And then good communities are the elements of a good C, for instance, which achieves a high $Q(G, C)$ score. Obviously, this definition of communities depends on the particular quality function Q used. A large number of quality functions have been proposed in [47]. The QACR protocol uses the Girvan-Newman modularity maximization [46], the most popular quality function. The $Q(G, C)$ is defined as in Eq. (5).

$$Q(G, C) = \sum_{k=1}^{K} \left(\frac{E_k^i}{2m} - \left(\frac{m_k}{2m} \right)^2 \right) \quad (5)$$

where, $m_k = \sum_{s_i \in c_k} \sum_{s_j \in S} A(s_i, s_j)$ is the number of edges of a cluster c_k and m is the number of edges of $G(S, E)$.

Once the clustering is completed, the BS calculates the rank of each cluster. The rank of a cluster is defined by the total RSS of N nodes from the BS and the total RSS of CMs in the cluster from the BS as in Eq. (6). A higher rank of a cluster indicates that the probability of the CMs nearer to the BS with good link quality (to some extent) is higher. The BS multicasts the individual clusters' information along with a *Hello_Msg4* to the CMs. A CM receives the message and lists the information in its cluster member table (CMT).

$$rank(c_k) = \frac{\sum_{i=1}^{N} |RSS(BS, s_i)|}{\sum_{j=1}^{n=|c_k|} |RSS(BS, s_j)|} \quad (6)$$

where, n is the number of CMs of cluster $c_k, k = 1, 2, 3, \ldots, K$, $RSS(BS, s_i)$ denotes the RSS of node s_i from the BS, and $RSS(BS, s_j)$ denotes the RSS of node $s_j \in c_k$ from the BS.

B. Cluster Head (CH) Election Phase

The number of CHs in a cluster is adaptive and depends on the different cases of a cluster state. The community detection algorithm does not consider the distance between nodes. Instead, it uses only the link information of the nodes. Therefore, some of the CMs can be out of R_n from other members in a large cluster. However, the CH election in each cluster depends on the following cases in two ways. The election is based on the rank of the CMs in all cases.

At the first round ($r = 1$), each CM calculates its rank and broadcasts the value of rank and residual energy by a *Node_Msg* within its R_{min}, where R_{min} is the distance of the farthest alive CM from the node in a cluster. A CM updates its CMT on received the message. The rank of a CM is defined by its residual energy, average RSS from CMs, and average link quality with CMs as in Eq. (7). The equation states that a CM having higher residual energy, higher average RSS, and higher average link quality has the higher rank. If the higher the rank, the better the option to achieve a better QoS during
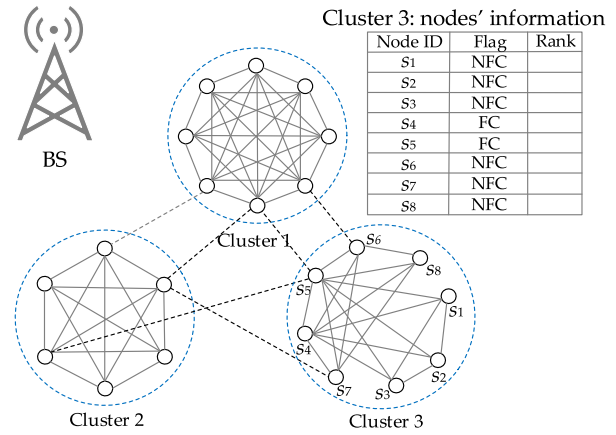

Fig. 3. An example of clustering of the nodes. Case 1: all cluster members (CMs) are within the communication range R_n from any other CM (defined by *flag* = 'FC'), see Clusters 1 and Cluster 2. Case 2: a set of nodes $\{s_1, s_2, s_3\}$ are out of R_n from some of the CMs (defined by *flag* = 'NFC') and a set $\{s_4, s_5\}$ belongs to an intersection region of two independent sets $\{s_1, \ldots, s_5\} \cap \{s_4, \ldots, s_8\}$, see Cluster 3.

intra-cluster communication.

$$rank(s_i) = \frac{E_{res}(s_i) \times \left(\frac{1}{n-1} \sum_{j \neq i}^{j=(n-1)} LQ(s_i, s_j) \right)}{\alpha \times E_{max} \times \left(\frac{1}{n-1} \sum_{j \neq i}^{j=(n-1)} |RSS(s_j, s_i)| \right)} \quad (7)$$

where, α is the weight factor between the value of $(0, 1]$, $E_{res}(s_i)$ is the residual energy of node s_i, E_{max} is the maximum initial energy of nodes S, n is the number of alive CMs belonging to the cluster of node s_i, $RSS(s_j, s_i)$ denotes the RSS of node s_i from node s_j, and $LQ(s_i, s_j)$ denotes the link quality between node s_i and node s_j. The factor α can be defined based on the QoS requirements of the intra-cluster communication.

Unlike the first round, from the second round ($r > 1$), a former CH of the previous round ($r - 1$) hands over the CH role to a next CH in a current round r. If the local data or the other information (i.e., rank and residual energy) has not been sent form a CM to the associate CH during the communication, the node is considered a missing node or a depleted node and will be eliminated from the remaining operations. In an exception, for example, if a *Schedule_Msg* has not been broadcasted from a tentative CH during the assigned time, the competition of the CH election in the cluster is similar to that described in the first round. Otherwise, a CH is elected by a former CH in a cluster recursively over the rounds.

Case 1: If all CMs in a cluster are within the R_n from any other CM and the nodes with *flag* = 'FC' (i.e., see Cluster 1 and 2 in Fig. 3), the higher ranked CM is elected as a CH. If there are multiple CMs have the same rank in a cluster, the residual energy of the nodes plays a role as the tie breaking metric, the higher residual energy obtained node is selected as a CH. An elected CH broadcasts the time division multiple access (TDMA) slots for the CMs by a *Schedule_Msg*. The number of slots is assigned according to the packet drop probability of a CH.

Node ID	Flag	Rank
s_1	NFC	2.3
s_2	NFC	2.6
s_3	NFC	2.5
s_4	FC	3.5
s_5	FC	3.4
s_6	NFC	2.4
s_7	NFC	2.7
s_8	NFC	2.8

(a)

Node ID	Flag	Rank
s_1	NFC	3.5
s_2	NFC	2.6
s_3	NFC	2.5
s_4	FC	2.1
s_5	FC	2.3
s_6	NFC	2.4
s_7	NFC	2.7
s_8	NFC	2.8

(b)

Node ID	Flag	Rank
s_1	NFC	2.3
s_2	NFC	2.6
s_3	NFC	2.5
s_4	FC	3.5
s_5	FC	3.4
s_6	NFC	2.4
s_7	NFC	3.8
s_8	NFC	2.8

(c)

Fig. 4. An illustration of rank and link status of the nodes in Cluster 3. The link status among cluster members (CMs) is defined by a *flag*. Table (a): a higher ranked node s_5 with *flag* = '*FC*' is belonged to an intersection region. Table (b): two independent sets are $\{s_1, s_2, s_3\}$ and $\{s_6, s_7, s_8\}$ where the nodes s_1 and s_8 with *flag* = '*NFC*' have the higher ranks in the sets, respectively. Table (c): a node s_4 in set $\{s_1, \ldots, s_5\}$ has the higher rank which is not common in the set $\{s_4, \ldots, s_8\}$.

Case 2: If some CMs in a cluster are not within the R_n from other CM and with *flag* = '*NFC*' (i.e., see Cluster 3 in Fig. 3), the number of CHs elected in the cluster varies according to the following three conditions.

Condition 1: A CM with *flag* = '*FC*' has the higher rank and belongs to an intersection region in a cluster and is elected as a CH (see Table (a) in Fig. 4). The rest of the process is same as in Case 1.

Condition 2: If multiple independent sets in a cluster exist, several CHs can be elected in the cluster. An example scenario can be seen in Cluster 3 in Fig. 3. The sets of nodes $\{s_1, s_2, s_3\}$ and $\{s_6, s_7, s_8\}$ are two independent sets after subtracting the subset $\{s_4, s_5\}$ from the set $\{s_1, \ldots, s_8\}$. Herein, the nodes with the higher ranks can belong in different independent sets (see Table (b) in Fig. 4), therefore, they are elected as the CHs intuitively. Each elected CH broadcasts a *Shcedule_Msg* within its R_{min}. The CMs in the intersection region receive the message(s) and decide the CH with the higher rank to join. A CM in the intersection region keeps the information of all elected CHs in the cluster.

Condition 3: A node may obtain the higher rank in a subset (i.e., s_4 in $\{s_4, s_5\} \in \{s_1, \ldots, s_5\}$). But the node may not be elected as a CH due to the possibility that another node has the higher rank in another subset (i.e., $\{s_4, \ldots, s_8\}$, see Table (c) in Fig. 4). Herein, the higher ranked node is elected as a CH and broadcasts a *Schedule_Msg* for the CMs in the cluster. The CMs within R_n receive the message and update their CMT. It may happen that a number of CMs can be out of R_n from an elected CH in a cluster; therefore, the message has not received by the nodes, it results in that the nodes are unaware of the elected CH. In this condition, an out-of-ranged node selects a CM with the higher rank and *flag* = '*FC*' from its CMT and sends a *Join_Msg* to the node to become a relay node towards the CH. On receiving the message, a relay node sends back a *Schedule_Msg* to the sender according to the time slot that has already been assigned by the CH. A relay node receives the local data from the node and forwards the packet during the allocated time slot.

Case 3: If there is only one node alive in a cluster, the node is elected as a CH without any competition. The data aggregation does not take place at the node.

C. Data Communication Phase

The data communication phase is divided to two subphases as follows.

1) Intra-Cluster Communication: During the allocated time for each node, a CM sends its local data to the respective CH using R_{min}, where R_{min} is the distance between a CM and the CH. A CH aggregates the received data into a packet called aggregated data. It is assumed that the sensory data of the nodes is highly correlated. If there is no exception (i.e., multiple CHs can be elected in a cluster, as in Condition 2 in Case 2), the values of rank and residual energy of a CM are piggybacked along with the local data sent to the CH. Otherwise, it also sends the values without data to other CHs during the assigned time slots by the CHs, accordingly.

2) Inter-Cluster Communication: Once the data aggregation is completed, inter-cluster communication can begin. During this time, the routing paths are constructed and aggregated data are forwarded towards the BS. In routing, a relay CH is selected based on the rank of the CHs. The rank of a CH is defined by its residual energy, RSS from the BS, and number of CMs. Each CH calculates its rank by using Eq. (8) and broadcasts it along with a *Route_Msg* within R_{max}. The CH resides the range, receives the message and updates its routing table (RT). The ranking function ensures that a CH having higher residual energy, higher RSS from the BS, and minimum number of CMs has the higher rank. The higher the rank, the better the option to achieve a better QoS during inter-cluster communication. If the BS is out of R_{max} from a CH, the CH selects one of the higher ranked CHs and belongs to a higher ranked cluster from RT as a relay node. When there are multiple CHs which have the same rank, the residual energy plays a role as the tie breaking metric, the higher residual energy obtained CH is selected as a relay node. The data received by a relay node from other CHs are not aggregated to send to the BS.

$$rank(CH_i) = \frac{E_{res}(CH_i)}{\beta \times n \times E_{max} \times |RSS(BS, CH_i)|} \quad (8)$$

where, β is the weight factor between the value of $(0,1]$, $E_{res}(CH_i)$ is the current residual energy of CH_i, n is the number of alive CMs belongs to the cluster of CH_i, E_{max} is the maximum initial energy of nodes S, and $RSS(BS, CH_i)$ denotes the RSS of CH_i from the BS. The factor β can be defined based on the QoS requirements of the inter-cluster communication.

If a data packet drops at a receiver, an automatic repeat request *ARR_Msg* is sent to the sender node. On receiving the message, the sender node retransmits the packet to the receiver. Each node stores a packet until (defined by the TH_{RD}) it is successfully received by the receiver. It may happen that a packet needs to be retransmitted a large number of times due to the bad quality of a link between nodes. To resolve this issue, there are many solutions that can be considered [48]. We consider a threshold TH_{RD} that can be defined by the link conditions and required QoS. If the number is exceeded, the packet is considered to be a lost packet.

TABLE II
SIMULATION PARAMETERS SETTING

Parameter	Value
Area of interest, M	(200×200) m^2
Initial energy, $[E_{min}, E_{max}]$	$[0.5, 1.5]$ J
Maximum transmission range of a node, R_{max}	180 m
Adaptive range, R_{adp}	$(\gamma = 0.4)\sqrt{(R_n{}^2/2)}$ m, $\gamma = (0, 1]$
Quality function, Q	$[0.5, 1]$
Cost in computation of rank and energy, E_{com}	5 nJ/bit/signal
Cost of data aggregation, E_{da}	5 nJ/bit/signal
Path loss exponent	Free space, 4
Noise distribution	Gaussian random noise, $X_\sigma = randn\ (\sigma = 0.1) + (\mu = 0)$
Packet drop probability of nodes	$[0, 0.68]$ defined by the signal-to-noise ratio
Transceiver circuitry, E_{elec}	50 nJ/bit
Threshold, (TH_{RC}, TH_{RD})	3
Cost of transmit amplifire, ε_{mp}	0.0013 pJ/bit/m^4, if $(d > d_0)$
Cost of transmit amplifier, ε_{fs}	10 pJ/bit/m^2, if $(d \leq d_0)$
Link quality, LQ	$good = 1$ and $bad = [0, 1)$
Weight factor, (α, β)	0.5

TABLE III
SIMULATION PARAMETERS SETTING FOR DIFFERENT NUMBER OF NODES

Parameter	Value
Location of the BS, (x, y)	$(100, 220)$ m
(N, R_n, R_{adp})	$\{(100, 120\ m, 34\ m), (200, 100\ m, 28\ m), (300, 80\ m, 22\ m), (400, 70\ m, 19\ m), (500, 50\ m, 14\ m)\}$
Location of the nodes	Random
(Local data packet, Aggregated data packet) size	$(2000, 4000)$ bits

V. PERFORMANCE EVALUATION

This section presents the simulation environment, parameters, performance and discussion of the protocols QACR, EADUC [25], EADC [29], CHP [37], and ECCR [38].

A. Simulation Setup

The simulations were conducted in MATLAB R2018a to evaluate the performance of the protocols. Two network topologies of Scenario 1: random with uniform and Scenario 2: random with non-uniform node distribution were considered. The nodes were heterogeneous in terms of initial energy. The weight factors α and β were set to 0.5 for both scenarios in QACR. The value of the factors for the other protocols was set accordingly. The R_n and R_{adp} were adjusted according to the node density and to avoid the frequent situation (nodes are not fully-connected in a cluster) that has been mentioned in Case 2 in Section IV, respectively. Table II shows the common parameters used in the simulations. Simulation results were averaged over 100 runs.

B. Simulation Results

The following performance metrics are evaluated and compared to the existing protocols for various network parameters. The parameters used in the simulations for various nodes N, packet sizes, and BS location are shown in Tables III, VII, and IX, respectively.

1) Packet Delivery Ratio (PDR): The PDR is the ratio of successfully received data at BS to the data sent. The PDR of ECCR, CHP, EADC, and EADUC are random due to the CH election and route selection policy of the protocols (i.e., the RSS and link quality are not considered), where a higher PDR is not guaranteed compared to QACR. In a random case, if the random noise is distributed over the network, the link conditions are varied among nodes. In this case, if N is constant, the packet drop and loss probability increase with increasing
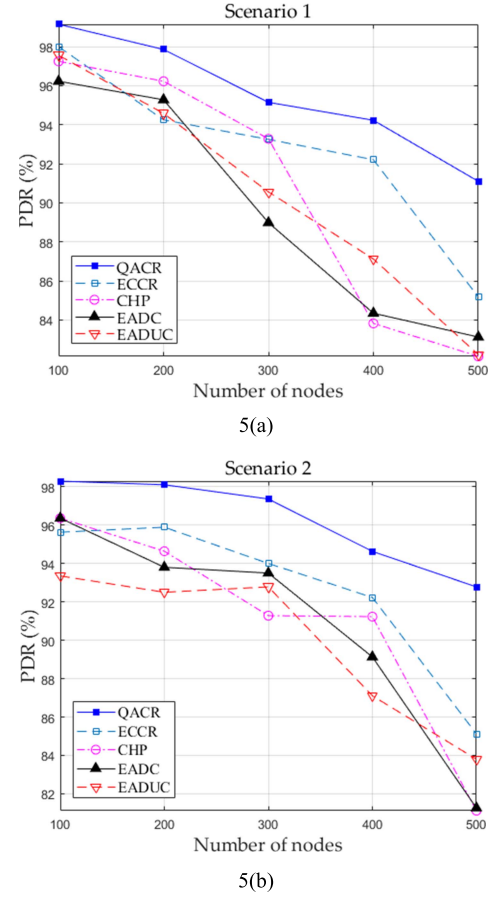

5(a)

5(b)

Fig. 5. Packet delivery ratio (PDR) for different number of sensor nodes under the scenarios; (a) Scenario 1: random with uniform node distribution, and (b) Scenario 2: random with non-uniform node distribution.

the number of elected CHs and the average number of hops towards the BS during communication. On the other hand, when the number of CHs is not varied significantly according to N, the metric also decreases with increasing the density of nodes due to the probability of the nodes associated with bad link quality to become the CHs increases. It is to be mentioned that, the average number of CHs in every round for QACR, ECCR, CHP, EADC, and EADUC were around 7, 14, 16, 5, and 10 respectively. The number of the CHs of QACR was higher than that of EADC and lower than that of ECCR, CHP, and EADUC. The CH election of the five protocols was periodic. The cluster formation of QACR and ECCR were static, whereas it was periodic in other protocols.

The RSS and link quality along with the residual energy of the nodes in CH election in QACR play an important role

TABLE IV
COMPARISON OF PACKET DELIVERY RATIO (PDR)
FOR THE NUMBER OF NODES

	Protocol	Avg. NL (round)	Avg. PDR (%)	QoS observation
Scenario 1	QACR	≈ 2256	95.50	QACR improves the PDR (2.93 ∼ 5.91) % of the other protocols in uniform node distribution
	ECCR	1987	92.57	
	CHP	1225	90.54	
	EADC	≈ 1346	89.59	
	EADUC	≈ 1654	90.41	
Scenario 2	QACR	2382	96.23	QACR improves the PDR (3.66 ∼ 6.32) % of the other protocols in non-uniform node distribution
	ECCR	≈ 2117	92.57	
	CHP	1591	90.93	
	EADC	1504	90.81	
	EADUC	1900	89.91	

to reduce the number of packets drop and loss. It is intuitive that when the distance between nodes is equal and the signal-to-noise ratio is higher, the RSS among them is higher. Nodes with higher RSS, the probability of packets drop is lower. Furthermore, if the noise is less and/or constant, the average RSS between nodes is higher as the average distance between them decreases, which ensures a minimum cost during communication. Fig. 5(a–b) provides the simulation results of PDR of the protocols. The results show that the PDR of the five protocols decrease with increasing the value of N. Because increasing of N results in increasing the number of packets in the network, the probability of packet collisions also increases. The comparison of the protocols concerning the PDR has been given in Table IV. The results depict that the QACR achieves a higher PDR than that of other protocols under the scenarios.

2) Latency: It measures the delay in the data reaching their destination across the network. We consider the delay proportional to the average number of hops in the data packets routing from CHs to the BS. The routing path selection of the protocols is according to their routing policy. The ECCR constructs a route among the relay nodes that are within $R_{\max}$ to each other, where the $R_{\max}$ is defined by the grid (cluster) size. According to the protocol, the value of $R_{\max}$ decreases with increases the value of N. The lower the value of the $R_{\max}$, the higher the number of hops required to construct a route. Unlike the ECCR, if the BS is within the communication range of all the nodes, the CHP always constructs a route from a CH to the BS using 2-hops. Each CH selects a dedicated relay node to send data to the BS. The EADC constructs a route with the goal of the load balancing during relay node selection which incurs to construct a route with a higher number of hops. On the other hand, the EADUC constructs a route based on the shortest distance where the number of hops is not taken into consideration. Although the constructed routes are shortest paths, some of the routes are constructed with a higher number of hops. In contrast, the routing policy of QACR utilizes the maximum capacity of the transmission range of a node to collect the relay nodes' information. The relay node selection criteria include the ranking metrics (the RSS in particular) and the rank of a cluster that ensures a reliable route with a minimum number of hops. Fig. 6(a–b) provides the simulation results of the average number of hops of the protocols. The results show when the
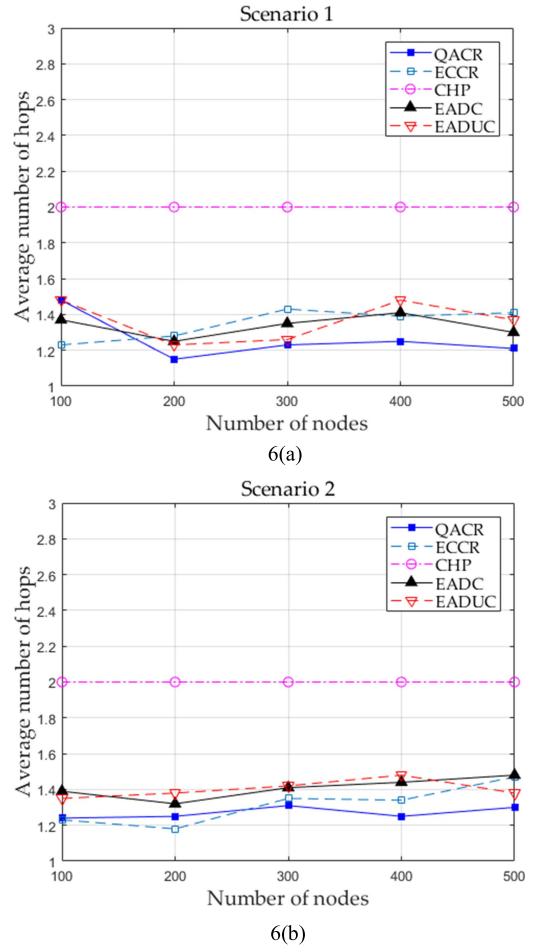

Fig. 6. Average number of hops of the data packets received at the BS for different number of sensor nodes under the scenarios; (a) Scenario 1: random with uniform node distribution, and (b) Scenario 2: random with non-uniform node distribution.

TABLE V
COMPARISON OF LATENCY FOR THE NUMBER OF NODES

	Protocol	Avg. NL (round)	Avg. no. of hops	QoS observation
Scenario 1	QACR	≈ 2256	1.25	QACR minimizes the avg. no. of hops (0.10 ∼ 0.75) of the other protocols in uniform node distribution
	ECCR	1987	1.35	
	CHP	1225	2	
	EADC	≈ 1346	1.33	
	EADUC	≈ 1654	1.36	
Scenario 2	QACR	2382	1.27	QACR minimizes the avg. no. of hops (0.04 ∼ 0.73) of the other protocols in non-uniform node distribution
	ECCR	≈ 2117	1.31	
	CHP	1591	2	
	EADC	1504	1.41	
	EADUC	1900	1.40	

node density is not higher (i.e., $N = 100$), the ECCR constructs routes with a minimum number of hops compared to QACR. The metric of ECCR increases with increases the value of N due to the number of clusters of ECCR increases compared to QACR (according to the clustering policy of the protocols). Table V compares the protocols in terms of the metric. These comparative results clearly demonstrate the QACR is able to construct routes using a minimum number of hops under the scenarios.
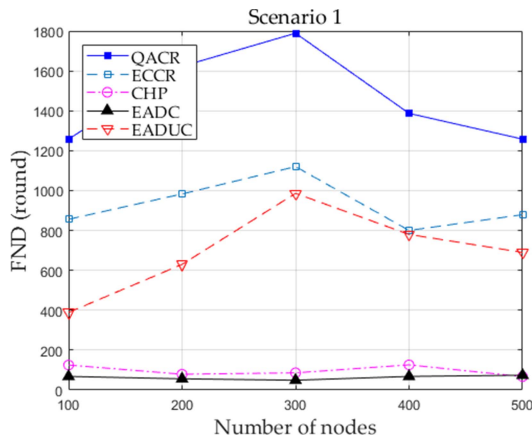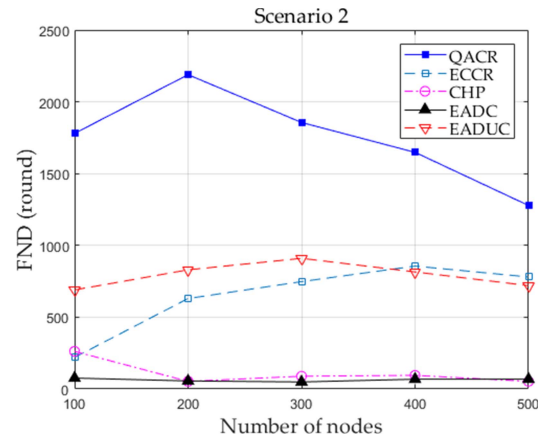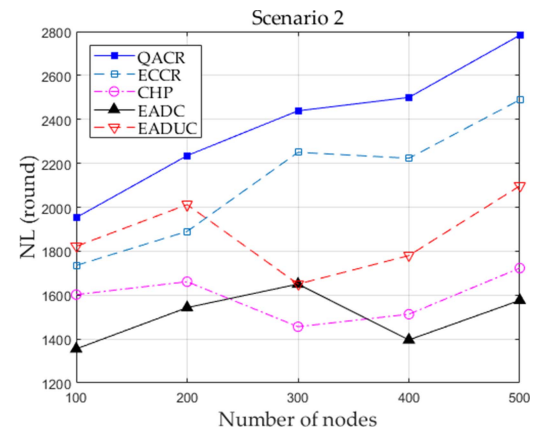
7(a)
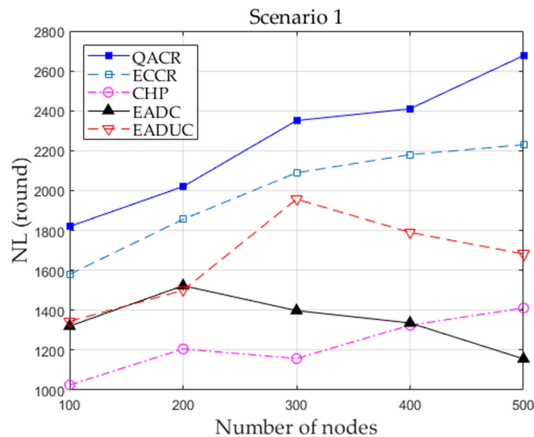
8(a)

7(b)

8(b)

Fig. 7.   Steady state and network lifetime (NL) for different number of sensor nodes under the Scenario 1: (a) First node dies (FND), and (b) Network lifetime (NL).

Fig. 8.   Steady state and network lifetime (NL) for different number of sensor nodes under the Scenario 2: (a) First node dies (FND), and (b) Network lifetime (NL).

3) Network Lifetime (NL): The steady state and NL are defined by the time the first node dies (FND) and a number of alive nodes (here, 60%), respectively. Unlike the clustering of QACR (community detection) and ECCR (static grid), the other protocols form cluster periodically based on the Voronoi cell which is defined by the transmission range of the advertisement message. Although the range of the cluster formation is equal, the clusters and the number of CMs are not distributed properly due to the random location of the elected CHs. Thus the energy consumption of the nodes is imbalanced. Meanwhile, some of the nodes are elected as CHs and participate as relay nodes towards the BS over the rounds repeatedly. In this regard, the nodes consume energy more frequently than other nodes and die early. This problem becomes severe when the nodes are distributed non-uniformly. In addition, the periodic cluster formation technique of CHP, EADC, and EADUC increases the control-message overhead throughout their NL and the nodes dissipate a significant amount of energy. In contrast, the clustering of QACR distributes the clusters and the number of CMs properly. It results in that the energy consumption among the nodes is distributed properly and the nodes save energy with the less control-message overhead. However, recall the CH election of the protocols, where the RSS and link condition are not considered, the nodes

dissipate an amount of energy due to the higher number of retransmissions of the dropped packets which has an impact on the metrics of the protocols. Unlike the other protocols, the nodes save energy regarding the less number of retransmissions by selecting the CHs associated with higher RSS and good link quality in QACR.

Figs. 7(a–b) and 8(a–b) provide the simulation results of steady state and NL of the protocols. As seen in the figures, although the metrics of QACR are higher with compared to the other protocols, the steady state of QACR tends to decrease after a certain N. This comes from that if a higher number of nodes is out of $R_{\max}$ from the BS, some of the nodes participate as relay nodes to relay the nodes' information towards the BS during the network initialization and consume a significant amount of energy. It results in that the nodes die early, which cause a shorter steady state of the network. However, the comparison between QACR and the other protocols in terms of the steady state and NL has been given in Table VI. The results show that the QACR improves the metrics than that of other protocols under the scenarios.

Figs. 9(a–b) and 10(a–b) provide the simulation results of steady state and NL of the protocols for different packet sizes. During the simulation, the average number of CHs in every

TABLE VI
COMPARISON OF STEADY STATE (FND) AND NETWORK
LIFETIME (NL) FOR THE NUMBER OF NODES

	Protocol	Avg. FND (round)	Avg. NL (round)	QoS observation
Scenario 1	QACR	1463	$\approx$ 2256	QACR enhances the FND (1.58 ~ 23.22) times and NL (1.14 ~ 1.84) times of the other protocols in uniform node distribution
	ECCR	$\approx$ 927	1987	
	CHP	$\approx$ 96	1225	
	EADC	63	$\approx$ 1346	
	EADUC	695	$\approx$ 1654	
Scenario 2	QACR	$\approx$ 1750	2382	QACR enhances the FND (2.21 ~ 27.80) times and NL (1.11 ~ 1.58) times of the other protocols in non-uniform node distribution
	ECCR	$\approx$ 646	$\approx$ 2117	
	CHP	$\approx$ 110	1591	
	EADC	$\approx$ 63	1504	
	EADUC	793	1900	

TABLE VII
SIMULATION PARAMETERS SETTING FOR DIFFERENT PACKET SIZES
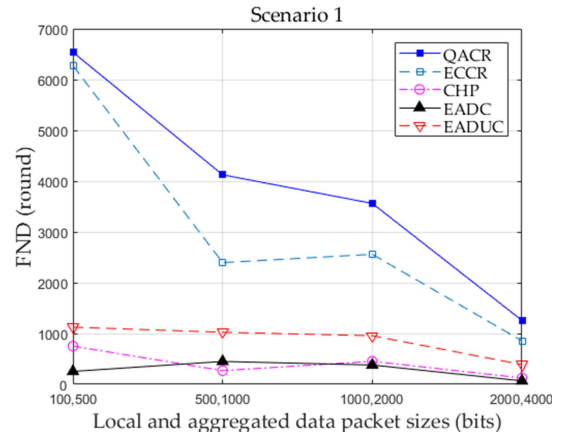
Parameter	Value
Location of the BS, (x, y)	(100, 220) m
(N, R_n, R_{adp})	(100, 120 m, 34 m)
Location of the nodes	Random and static
(Local data packet, Aggregated data packet) size	{(100, 500), (500, 1000), (1000, 2000), (2000, 4000)} bits

TABLE VIII
COMPARISON OF STEADY STATE (FND) AND NETWORK
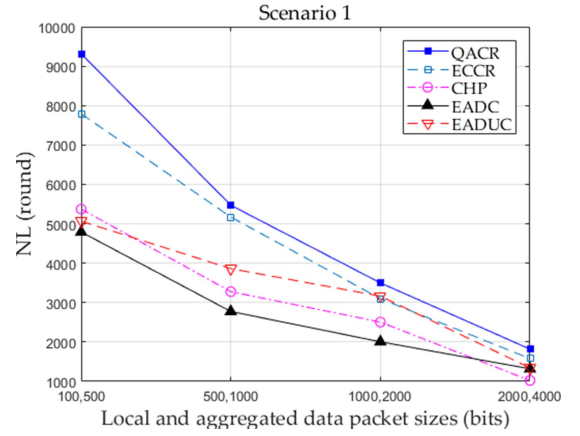LIFETIME (NL) FOR THE PACKET SIZES

	Protocol	Avg. FND (round)	Avg. NL (round)	QoS observation
Scenario 1	QACR	3871	$\approx$ 5026	QACR enhances the FND (1.28 ~ 13.44) times and NL (1.14 ~ 1.85) times of the other protocols in uniform node distribution
	ECCR	$\approx$ 3021	$\approx$ 4410	
	CHP	$\approx$ 400	$\approx$ 3047	
	EADC	288	2723	
	EADUC	$\approx$ 873	$\approx$ 3359	
Scenario 2	QACR	4109	$\approx$ 5099	QACR enhances the FND (2.76 ~ 21.63) times and NL (1.06 ~ 1.77) times of the other protocols in non-uniform node distribution
	ECCR	762	$\approx$ 4802	
	CHP	$\approx$ 688	3617	
	EADC	$\approx$ 190	$\approx$ 2884	
	EADUC	1488	$\approx$ 3722	

Fig. 9. Steady state and network lifetime (NL) for different packet sizes under the Scenario 1: (a) First node dies (FND), and (b) Network lifetime (NL).

TABLE IX
SIMULATION PARAMETERS SETTING FOR BS AT CENTER

Parameter	Value
Location of the BS, (x, y)	(100, 100) m
(N, R_n, R_{adp})	(100, 120 m, 34 m)
Location of the nodes	Random
(Local data packet, Aggregated data packet) size	(2000, 4000) bits

round for QACR, ECCR, CHP, EADC, and EADUC were around 5, 6, 7, 4, and 9, respectively. If N, network area and location of the BS are constant, the packet size does not have a greater impact on the performance of the protocols concerning the PDR and latency, but the NL of the protocols are varied significantly. The results depict that the NL of the five protocols decrease with increasing the packet size as the transmission and processing cost increase. The comparison between QACR and the other protocols regarding the steady state and NL has been given in Table VIII. The results show that the QACR outperforms the protocols in terms of the metrics under the scenarios.

Fig. 11(a–b) provides a randomly selected simulation result of steady state and NL of the protocols for the location of the BS at center (BS is within the maximum range of all the nodes). During the simulation, the average number of elected CHs in every round for QACR, ECCR, CHP, EADC, and EADUC were around 4, 7, 3, 3, and 6, respectively. If N, network area and packet size are constant, the NL of the five protocols increase with decreasing the average distance of the nodes from the BS. The comparison between QACR and other protocols in terms of the steady state and NL has

been given in Table X. The results show that the QACR improves the metrics compared to the protocols under the scenarios.

C. Discussion

Before drawing conclusions, we first highlight the key points in our proposition. Then we further discuss some design limitations and extension of QACR for future improvement. Based on the performance analysis, it is clear that QACR has overall superior performance in comparison with other protocols under various scenarios. The performance supremacy comes from several factors shown in Table XI are summarized as follows:

1) The metrics are used in clustering, CH election, and route selection in QACR to utilize the network resources
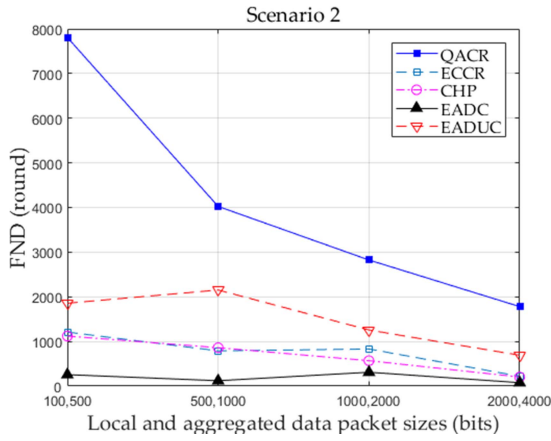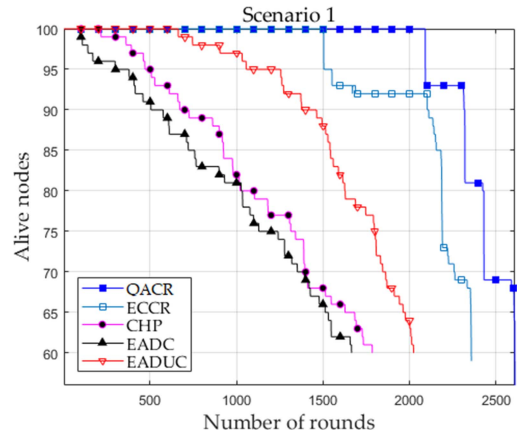
10(a)
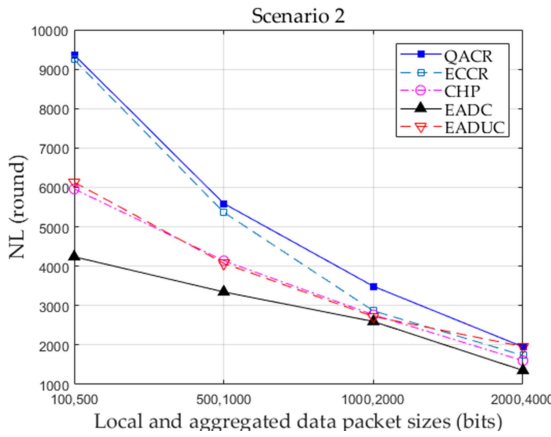
10(b)

Fig. 10. Steady state and network lifetime (NL) for different packet sizes under the Scenario 2: (a) First node dies (FND), and (b) Network lifetime (NL).

11(a)

11(b)

Fig. 11. Steady state and network lifetime (NL) for BS at center under the Scenario 1: and Scenario 2: (a) First node dies (FND), and (b) Network lifetime (NL).

TABLE X
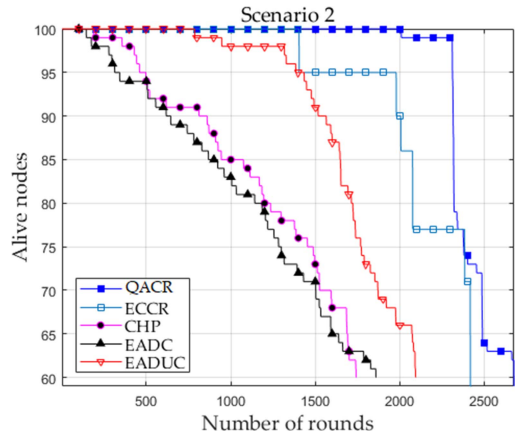COMPARISON OF STEADY STATE (FND) AND NETWORK
LIFETIME (NL) FOR THE BS AT CENTER

	Protocol	Avg. FND (round)	Avg. NL (round)	QoS observation
Scenario 1	QACR	≈ 1958	2569	QACR enhances the FND (1.03 ~ 24.20) times and NL (1.16 ~ 1.62) times of the other in uniform node distribution
	ECCR	1520	≈ 2215	
	CHP	≈ 193	1689	
	EADC	≈ 81	≈ 1584	
	EADUC	673	≈ 1893	
Scenario 2	QACR	≈ 2184	2658	QACR enhances the FND (1.53 ~ 20.13) times and NL (1.13 ~ 1.55) times of the protocols in non-uniform node distribution
	ECCR	≈ 1423	≈ 2110	
	CHP	154	≈ 1720	
	EADC	120	≈ 1920	
	EADUC	≈ 810	2089	

effectively, then to achieve a better QoS of the network. Residual energy metric is to select a node having high energy, RSS to measure the approximate distance and link condition (to some extent), link quality to provide reliability, and number of CMs to select a relay CH of minimum CMs.

2) The adopted clustering technique distributes the nodes properly with available nodes' connectivity in the clusters that reduces the number of packets loss and balances the energy consumption of the nodes.

3) The static clustering and CH role handover technique similar to ECCR reduces the control-message overhead and energy dissipation of the nodes for the periodic CH election and cluster formation.

The criteria of CH election and relay node selection include the most influential factors of the QoS related with the data communication that achieves a higher PDR, lower latency and prolonged NL.

While QACR proved its efficiency in the simulation tests, there are still some aspects that need further analysis. In our proposed CH election and relay node selection, the value of α and β were constant throughout the network lifetime. The value of the factors could be optimized and dynamic according to the network states using some latest techniques such as machine learning. However, it should be reasonable for resource constrained devices in LLNs. The QACR was described by which the link quality and the packet drop probability of the nodes are defined by the Gaussian random noise in static. The link condition can be dynamic and varied over time. This might have an impact on the performance of the protocols. In view of the significance and dynamic usability of this protocol for LLNs, we leave this issue for future improvement.

TABLE XI
TECHNIQUES USED IN QACR AND EXISTING PROTOCOLS

Protocol	Clustering technique	CH election technique	CH election metrics				CH re-election technique (Handover)
			Residual energy	No. of neighbors	Distance	Link quality	
EADUC [25]	Voronoi cell	Distributed	Yes	No	No	No	No
EADC [29]	Voronoi cell	Distributed	Yes	No	No	No	No
CHP [37]	Voronoi cell	Centralized	Yes	No	to BS	No	No
ECCR [38]	Optimum number of grids	Distributed	Yes	Yes	to nodes	No	Yes
QACR	Community detection	Distributed	Yes	Yes	to nodes	Yes	Yes

VI. CONCLUSION

This paper presents a QoS-aware data collection protocol for LLNs in Fog-enabled IoT. It applies an effective technique in clustering, CH election, CH re-election and route selection for the network. The clustering adopts the community detection algorithm which is considered more resilient than other clustering techniques. The CH election and route selection are defined by a set of metrics that reduces the packets loss, number of hops in data routing and balances the energy consumption of the nodes. As a result, the network can sustain a prolonged lifetime with a higher packet delivery ratio and lower latency. Two different network topologies of a LLN with uniform and non-uniform node distribution were considered and tested over the described experimental setup. The results validate that QACR protocol improves the QoS of the LLN as compared to the existing protocols under various scenarios.

REFERENCES

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.

[2] F. Y. Okay and S. Ozdemir, "Routing in fog-enabled IoT platforms: A survey and an SDN-based solution," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 4871–4889, Dec. 2018.

[3] H. F. Atlam, R. J. Walters, and G. B. Wills, "Fog computing and the Internet of Things: A review," *Big Data Cogn. Comput.*, vol. 2, no. 2, pp. 1–18, 2018.

[4] J. N. Al-Karaki and A. Gawanmeh, "The optimal deployment, coverage, and connectivity problems in wireless sensor networks: Revisited," *IEEE Access*, vol. 5, pp. 18051–18065, 2017.

[5] M. T. Lazarescu, "Design of a WSN platform for long-term environmental monitoring for IoT applications," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 3, no. 1, pp. 45–54, Mar. 2013.

[6] A. Sarkar and T. S. Murugan, "Routing protocols for wireless sensor networks: What the literature says?" *Alexandria Eng. J.*, vol. 55, no. 4, pp. 3173–3183, 2016.

[7] N. Javaid, M. R. Jafri, Z. A. Khan, N. Alrajeh, M. Imran, and A. Vasilakos, "Chain-based communication in cylindrical underwater wireless sensor networks," *Sensors*, vol. 15, no. 2, pp. 3625–3649, 2015.

[8] I. Lamprinos, M. Charalambides, and M. Chouchoulis, "Greenhouse monitoring system based on a wireless sensor network," in *Proc. 2nd Int. Electron. Conf. Sensors Appl.*, Nov. 2015, pp. 15–30.

[9] C. Renner and P. A. T. Nguyen, "A system for efficient dissemination of weather forecasts for sustainable solar-powered sensors," in *Proc. Sustain. Internet ICT Sustain.*, Madrid, Spain, Apr. 2015, pp. 14–15.

[10] X. Li, A. Liu, N. N. Xiong, Z. Zeng, and Z. Cai, "Adaptive aggregation routing to reduce delay for multi-layer wireless sensor networks," *Sensors*, vol. 18, no. 4, pp. 1–28, 2018.

[11] C.-I. Wu, H.-Y. Kung, C.-H. Chen, and L.-C. Kuo, "An intelligent slope disaster prediction and monitoring system based on WSN and ANP," *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4554–4562, 2014.

[12] N. A. Pantazis, S. A. Nikolidakis, and D. D. Vergados, "Energy-efficient routing protocols in wireless sensor networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 2, pp. 551–591, 2nd Quart., 2013.

[13] S. Dehghani, M. Pourzaferani, and B. Barekatain, "Comparison on energy-efficient cluster based routing algorithms in wireless sensor network," *Procedia Comput. Sci.*, vol. 72, pp. 535–542, 2015. doi: 10.1016/j.procs.2015.12.161.

[14] Y. Zhou, X. Wang, T. Wang, B. Liu, and W. Sun, "Fault-tolerant multipath routing protocol for WSN based on HEED," *Int. J. Sensor Netw.*, vol. 20, no. 1, pp. 37–45, 2016.

[15] Z. Tang, K. Liu, J. Xiao, L. Yang, and Z. Xiao, "A parallel *k*-means clustering algorithm based on redundance elimination and extreme points optimization employing MapReduce," *Concurrency Comput. Pract. Exp.*, vol. 29, no. 20, pp. 1–18, 2017.

[16] X. Yin, K. Zhang, B. Li, A. K. Sangaiah, and J. Wang, "A task allocation strategy for complex applications in the heterogeneous cluster-based wireless sensor networks," *Int. J. Distrib. Sensor Netw.*, vol. 14, no. 8, pp. 1–15, 2018.

[17] T. Wang, L. Wen, Y. Zhou, and J. Zhang, "A energy balanced routing scheme in wireless sensor networks based on non-uniform layered clustering," *Int. J. Sensor Netw.*, vol. 27, no. 4, pp. 239–249, 2018.

[18] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. 33rd Annu. Hawaii Int. Conf. Syst. Sci.*, Jan. 2000, pp. 1–10.

[19] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrisnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 660–670, Oct. 2002.

[20] L. Yan, Z. X. Huang, and L. Y. Zhong, "Algorithm of cluster head multi-hops based on LEACH," *Comput. Eng. Design*, vol. 28, no. 17, pp. 4158–4160, 2008.

[21] L. Ming, C. J. Nong, C. G. Han, C. L. Jun, W. X. Min, and G. H. Gang, "EADEEG: An energy-aware data gathering protocol for wireless sensor networks," *J. Softw.*, vol. 18, no. 5, pp. 1092–1109, 2007.

[22] R. Bista and A. Shakya, "Randomized distance based clustering algorithm for energy efficient wireless sensor networks," *Adv. Comput. Commun. Technol.*, vol. 452, pp. 377–389, Jun. 2016.

[23] N. Javaid *et al.*, "An energy-efficient distributed clustering algorithm for heterogeneous WSNs," *EUROSHIP J. Wireless Commun. Netw.*, vol. 151, pp. 1–11, Jun. 2015.

[24] S. K. Singh, P. Kumar, and J. P. Singh, "A survey on successors of LEACH protocol," *IEEE Access*, vol. 5, pp. 4298–4328, 2017.

[25] J. Yu, Y. Qi, G. Wang, Q. Guo, and X. Gu, "An energy-aware distributed unequal clustering protocol for wireless sensor networks," *Int. J. Distrib. Sensor Netw.*, vol. 7, no. 1, pp. 1–8, 2011.

[26] Z. Qin, C. Ma, L. Wang, J. Xu, and B. Lu, "An overlapping clustering approach for routing in wireless sensor networks," *Int. J. Distrib. Sensors Netw.*, vol. 9, no. 3, pp. 1–11, 2013.

[27] O. Zytoune, Y. Fakhri, and D. Aboutajdine, "Lifetime maximization algorithm in wireless sensor network," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 6, no. 3, pp. 140–149, 2010.

[28] S.-S. Wang and Z.-P. Chen, "LCM: A link-aware clustering mechanism for energy-efficient routing in wireless sensor networks," *IEEE Sensors J.*, vol. 13, no. 2, pp. 728–736, Feb. 2013.

[29] J. Yu, Y. Qi, G. Wang, and X. Gu, "A cluster-based routing protocol for wireless sensor networks with nonuniform node distribution," *Int. J. Electron. Commun.*, vol. 66, no. 1, pp. 54–61, 2012.

[30] M. Sabet and H. R. Naji, "A decentralized energy efficient hierarchical cluster-based routing algorithm for wireless sensor networks," *Int. J. Electron. Commun.*, vol. 69, no. 5, pp. 790–799, 2015.

[31] Y. Yao, Q. Cao, and A. V. Vasilakos, "EDAL: An energy-efficient, delay-aware, and lifetime-balancing data collection for heterogeneous wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 23, no. 3, pp. 810–823, Jun. 2015.

[32] X. Wang, Z. Zhao, Y. Xia, and H. Zhang, "Compressed sensing for efficient random routing in multi-hop wireless sensor networks," in *Proc. IEEE Globecom Workshops*, Miami, FL, USA, Dec. 2010, pp. 266–271.

[33] A. Liu, Z. Zheng, C. Zhang, Z. Chen, and X. Shen, "Secure and energy efficient disjoint multipath routing for WSNs," *IEEE Trans. Veh. Technol.*, vol. 61, no. 7, pp. 3255–3265, Sep. 2012.

[34] Z. Xu, L. Chen, L. Cao, T. Liu, D. Yang, and C. Chen, "DARC: A distributed and adaptive routing protocol in cluster-based wireless sensor networks," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 12, pp. 1–14, Dec. 2015.

[35] L. Malathi, R. K. Gnanamurthy, and K. Chandrasekaran, "Energy efficient data collection through hybrid unequal clustering for wireless sensor networks," *Comput. Elect. Eng.*, vol. 48, pp. 358–370, Nov. 2015.

[36] V. Gupta and R. Pandey, "An improved energy aware distributed unequal clustering protocol for heterogeneous wireless sensor networks," *Int. J. Eng. Sci. Technol.*, vol. 19, no. 2, pp. 1050–1058, 2016.

[37] Y. Zhou, N. Wang, and W. Xiang, "Clustering hierarchy protocol in wireless sensor networks using an improved PSO algorithm," *IEEE Access*, vol. 5, pp. 2241–2253, 2016.

[38] A. S. M. S. Hosen and G. H. Cho, "An energy centric cluster-based routing protocol for wireless sensor networks," *Sensors*, vol. 18, no. 5, pp. 1–17, 2018.

[39] M. V. Kumaramangalam, K. Adiyappatham, and C. Kandasamy, "Zone-based routing protocol for wireless sensor networks," *Int. Scholarly Res. Notices*, vol. 2014, pp. 1–9, Nov. 2014.

[40] S. B. Gowda, C. Puttamadappa, H. S. Mruthyunjaya, and N. V. Babu, "Sector based multi-hop clustering protocol for wireless sensor networks," *Int. J. Comput. Appl.*, vol. 43, no. 13, pp. 32–38, 2012.

[41] X. Chen and N. C. Rowe, "Saving energy by adjusting transmission power in wireless sensor networks," in *Proc. IEEE Glob. Telecommun. Conf.*, Kathmandu, Nepal, Dec. 2011, pp. 1–5.

[42] S. Lin *et al.*, "ATPC: Adaptive transmission power control for wireless sensor networks," *ACM Trans. Sensor Netw.*, vol. 12, no. 1, pp. 1–31, 2016.

[43] A. Ahmad, N. Javaid, M. Imran, M. Guizani, and A. A. Alhamed, "An advanced energy consumption model for terrestrial wireless sensor networks," in *Proc. Int. Conf. Wireless Commun. Mobile Comput.*, Paphos, Cyprus, Sep. 2016, pp. 790–793.

[44] B. Li, W. Cui, and B. Wang, "A robust wireless sensor network localization algorithm in mixed LOS/NLOS scenario," *Sensors*, vol. 15, no. 9, pp. 23536–23553, 2015.

[45] D. S. J. D. Couto, D. Aguayo, J. Bicket, and R. Morris, "A high-throughput path metric for multi-hop wireless routing," *Wireless Netw.*, vol. 11, no. 4, pp. 419–434, 2005.

[46] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, Feb. 2004, Art. no. 026113.

[47] M. A. Javed, M. S. Younis, S. Latif, J. Qadir, and A. Baig, "Community detection in networks: A multidisciplinary review," *J. Netw. Comput. Appl.*, vol. 108, pp. 87–111, Apr. 2018.

[48] C. Phillips, D. Sicker, and D. Grunwald, "A survey of wireless path loss prediction and coverage mapping methods," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 255–270, 1st Quart., 2013.

Saurabh Singh received the bachelor's degree from Uttar Pradesh Technical University, the master's degree in information security from Thapar University, India, and the Ph.D. degree in computer science and engineering from Chonbuk National University. He is a Post-Doctoral Researcher with the School of Computer, Information and Communication Engineering, Kunsan National University, South Korea. He holds a strong academic record. He has published papers in the journals IEEE ACCESS, *IEEE Magazine*, the IEEE INTERNET OF THINGS JOURNAL, the *Journal of Network and Computer Applications*, the *Journal of Official Statistics*, and *Sustainability*. His research interests include cloud security, IoT, SNS, smart home, blockchain, wireless sensor networks, ubiquitous computing security, and cryptography. He is also a Reviewer of the IEEE INTERNET OF THINGS JOURNAL, *Computers & Electrical Engineering*, the *International Journal of Communication Systems*, *Future Generation Computing Systems*, and the *Journal of Human-Centric Computing and Information Science*.

Pradip Kumar Sharma received the master's degree in computer science from Thapar University, India, in 2014, and the Ph.D. degree in CSE from the Seoul National University of Science and Technology, South Korea. He was a Software Engineer with MAQ Software, India. He is a Post-Doctoral Research Fellow with the Department of Multimedia Engineering, Dongguk University, South Korea. He worked on a variety of projects, proficient in building large-scale complex data warehouses, OLAP models, and reporting solutions that meet business objectives and align IT with business. His current research interests are focused on the areas of ubiquitous computing, security, cloud computing, SDN, SNS, blockchain, and IoT.

Md. Sazzadur Rahman received the B.Sc. and M.S. degrees in applied physics, electronics and communication engineering from the University of Dhaka, Bangladesh, in 2005 and 2006, respectively, and the Ph.D. degree in material science from Kyushu University, Japan, in 2015. He was a Faculty Member with the Faculty of Computer Science and Engineering, Hajee Mohammad Danesh Science and Technology University from May 2009 to November 2018. Since November 2018, he has been with the Institute of Information Technology, Jahangirnagar University, Dhaka, Bangladesh, where he is currently as an Assistant Professor. His research interests include material science for computer application, surface science, ubiquitous computing and WSNs, blockchain, and IoT.

A. S. M. Sanwar Hosen received the B.S. degree in computer science and engineering (CSE) from East West University, Dhaka, Bangladesh, in 2010, and the M.S. and Ph.D. degrees in CSE from Chonbuk National University (CBNU), Jeonju, South Korea, in 2013 and 2017, respectively. He was a Post-Doctoral Researcher with the School of Computer, Information and Communication Engineering, Kunsan National University, and the Division of CSE, CBNU, from March 2017 to August 2019. He is currently a Research Assistant Professor with the Division of CSE, CBNU. He has published several articles in journals and international conferences. His research interests are in wireless sensor networks, Internet of Things, network security, data distribution services, artificial intelligence, blockchain, and green IT. He serves as a Reviewer of several journals, including the IEEE INTERNET OF THINGS JOURNAL, the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE ACCESS, *Sensors*, *Energies*, *Electronics*, and *Future Generation Computing Systems*.

In-Ho Ra received the Ph.D. degree in computer engineering from Chung-Ang University, Seoul, South Korea, in 1995. He has been with the School of Computer, Information and Communication, Kunsan National University, where he is currently a Professor. From February 2007 to August 2008, he was a Visiting Scholar with the University of South Florida, Tampa, FL, USA. His major research interests include wireless ad hoc and sensor network, blockchain, IoT, PS-LTE, and microgrid.

Gi Hwan Cho received the B.S. degree in computer science from Chonnam University, Gwangju, South Korea, in 1985, the M.S. degree in computer science from Seoul National University, Seoul, South Korea, in 1987, and the Ph.D. degree in computing science from the University of Newcastle, Newcastle upon Tyne, U.K., in 1996. He was with Electronics and Telecommunications Research Institute, Daejeon, South Korea. From September 1987 to August 1997, he was a Senior Member of Technical Staff with the Department of Computer Science, Mokpo National University, Mokpo, South Korea. From September 1997 to February 1999, he was a full-time Lecturer. In March 1999, he joined the Division of Computer Science and Engineering, Chonbuk National University, Jeonju, South Korea, where he is currently serving as a Professor. His current research interests include mobile computing, computer communication, security on wireless networks, WSNs, and distributed computing system.

Deepak Puthal (SM'19) was a Lecturer with the University of Technology Sydney, Australia. He is a Lecturer (Assistant Professor) with the School of Computing, Newcastle University, U.K. His research spans several areas in cyber security, blockchain, Internet of Things, and edge/fog computing. He was a recipient of the several recognitions and best paper awards from IEEE. He serves on the editorial boards of top quality international journals, including the IEEE TRANSACTIONS ON BIG DATA, *IEEE Consumer Electronics Magazine*, *Computers & Electrical Engineering* (Elsevier), the *International Journal of Communication Systems* (Wiley), and *Internet Technology Letters* (Wiley).