Received December 27, 2020, accepted January 10, 2021, date of publication January 13, 2021, date of current version January 20, 2021.
Digital Object Identifier 10.1109/ACCESS.2021.3051360

Q-Learning-Based Data-Aggregation-Aware Energy-Efficient Routing Protocol for Wireless Sensor Networks

WAN-KYU YUN AND SANG-JO YOO, (Member, IEEE)
Department of Information and Communication Engineering, Inha University, Incheon 402-751, South Korea
Corresponding author: Sang-Jo Yoo (sjyoo@inha.ac.kr)

This work was supported by the Inha University Research Grant.

ABSTRACT The energy consumption of the routing protocol can affect the lifetime of a wireless sensor network (WSN) because tiny sensor nodes are usually difficult to recharge after they are deployed. Generally, to save energy, data aggregation is used to minimize and/or eliminate data redundancy at each node and reduce the amount of the overall data transmitted in a WSN. Furthermore, energy-efficient routing is widely used to determine the optimal path from the source to the destination, while avoiding the energy-short nodes, to save energy for relaying the sensed data. In most conventional approaches, data aggregation and routing path selection are considered separately. In this study, we consider the degrees of the possible data aggregation of neighbor nodes when a node needs to determine the routing path. We propose a novel Q-learning-based data-aggregation-aware energy-efficient routing algorithm. The proposed algorithm uses reinforcement learning to maximize the rewards, defined in terms of the efficiency of the sensor-type-dependent data aggregation, communication energy and node residual energy, at each sensor node to obtain an optimal path. We used sensor-type-dependent aggregation rewards. Finally, we performed simulations to evaluate the performance of the proposed routing method and compared it with that of the conventional energy-aware routing algorithms. Our results indicate that the proposed protocol can successfully reduce the amount of data and extend the lifetime of the WSN.

INDEX TERMS Wireless sensor networks, routing, data aggregation, Q-learning, network lifetime.

I. INTRODUCTION

A wireless sensor network (WSN) can be defined as a self-configured and infrastructure-less wireless network used to monitor and record the physical conditions of an environment and store the collected data at a central location. WSNs have received considerable attention for multiple types of applications because of their low cost, small size and applicability in diverse fields such as healthcare, military and underwater monitoring [1]. Recently, the device, network and data management technologies for WSNs have been extended to other fields such as smart factories, where sensor nodes are deployed to collect data on products and machines for smart factory operations. In smart cities, WSNs can be deployed to create an efficient service delivery platform for public and municipal workers and to manage the city resources efficiently [2], [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Hongwei Du.

In a WSN, many sensor nodes are deployed over a wide area to collect observation data and send them to a sink (or server). Therefore, multi-hop transmission is required to deliver the collected data successfully to the sink located beyond the transmission range of the source sensor node. This requires a collecting sensor node to calculate the optimal route to the sink. Energy efficiency is a primary challenge to the successful application of WSNs because nodes have limited energy and cannot be recharged easily after they have been deployed. Furthermore, because energy is mostly consumed by the radio device, an energy-efficient design of the routing algorithm for communication is essential. Most of the ongoing research on energy-aware routing has two objectives: to minimize the overall energy consumption on the routing path and maintain even residual energy levels. Because the overall energy consumption depends on the distance between nodes and the number of intermediate nodes, the minimum hop count path or shortest distance path is generally used for WSN routing. The residual energy level of each node or power drain rate is also considered to avoid path disconnection and network partition. These measures can prolong the network lifetime because energy is dissipated more equally among all nodes [4], [5].

Because the data being collected by multiple sensors in a given area are based on common phenomena, there is likely to be some redundancy in the source data. Data aggregation as a form of ''in-network-processing'' in WSNs is widely used to collect data in an energy-efficient manner by eliminating redundancy and minimizing the number of transmissions or data size. In many WSN applications, the actual measured raw data at each sensor node may not need to be delivered in the exact same form to the sink. The raw data can be abstracted or compressed in networks. Depending on the monitoring purposes of applications, various aggregation techniques can be used, such as abstracting as {mean, variance}, maximum value, minimum value, lossy compression, feature domain reduction and data prediction. The efficiency of data aggregation increases when the correlation among the data collected by various sensors is high [6], [7].

Various machine learning technologies have been used to effectively capture the dynamic features such as node topology changes, restricted energy conditions, event detection and communication costs of WSNs for their energy-efficient operation. Among them, reinforcement learning (RL) is particularly suitable for problems that include a long- versus short-term reward trade-off. It provides a framework for a system to learn from its previous interactions with its environment and to select its actions efficiently in the future. RL-based routing protocols can determine the optimal path as an adaptive method for complex network conditions and quality of service requirements [8]–[10].

Most previous studies on energy-efficient routing path selection typically consider communication energy with hop counts and the distance to the sink node to reduce the overall network-wide energy consumption and/or residual energy level at each sensor node to distribute the energy burden equally. However, distributing the possible routes to reduce the overhead of specific sensor nodes may conflict with the objective of minimizing the network-wide energy consumption. Notably, the optimization goals do not consider the possibility of data aggregation through the path. Furthermore, data aggregation and routing path selection are considered separately in conventional approaches [11]–[14].

In this article, we propose an RL-based energy-aware routing algorithm for obtaining a global optimum path to minimize the overall energy consumption and prolong the lifetime of the WSN. We define the degrees of the possible data aggregation of neighbor nodes when a node needs to determine the routing path. Because data from various sensor types (e.g., vibration measurement sensor and temperature sensor) may not show strong correlation, they cannot be aggregated together. Therefore, we define sensor-type-dependent aggregation rewards. We propose a novel Q-learning-based data-aggregation-aware energy-efficient routing (Q-DAEER) algorithm, in which each sensor node reinforces to determine the optimum path that can maximize the rewards by considering the sensor-type-dependent data aggregation level of the neighbor node, the residual energy, communication cost with distance and hop count to the sink. In this way, the sensor nodes can determine the optimum next hop node using their updated Q-values based on the rewards.

This article is organized as follows: In Section II, we review the existing energy-aware routing protocols for the WSN. In Section III, we present our proposed system model for WSN routing. In Section IV, we discuss Q-DAEER algorithm. We present the simulation results in Section V and conclude this article in Section VI.

II. RELATED WORK

Routing is essential in WSNs to support reliable data transfer, achieve low latency and provide energy-efficient operation. Wireless communications consume significant amount of power for transmitting sensed data from sensor nodes to sink nodes. However, the power consumption has become a limiting factor because most sensor nodes are powered by batteries. Sensor nodes used in wireless networks have limited computational capability and cannot have full information about networks so that it is very difficult for nodes to calculate the optimum route to the destination quickly. Even when a node is able to obtain the optimum routing path, the path may not remain optimum over time owing to various types of changes in the sensing environment, for example, the node movement, instable wireless channel condition and dynamic energy status of sensor nodes. Conventional ad hoc routing protocols can be classified into proactive and reactive protocols [15]. In proactive routing, routes are computed even when they are not needed and stored in a routing table at every node. Therefore, the routing table maintenance overhead is large and limits the scalability of this routing protocol. In reactive routing, routes are computed only when they are needed, and sensor nodes store routes only for their neighbors. However, this protocol may increase latency for sensed data delivery. To overcome these problems, many studies on finding the optimum routing path with low energy consumption are underway.

Mohemed *et al.* [16] addressed the hole problem in WSNs using two distributed, energy-efficient and connectivity-aware routing protocols. They used two different protocols in local and global environments. This technique can decrease the overhead of topology reformation and prolong the network lifetime. Razaque *et al.* [17] presented the combined protocol of low-energy adaptive-clustering hierarchy (LEACH) and power-efficient gathering in sensor information systems (PEGASIS), named P-LEACH. This protocol can improve the performance by considering the limitation of cluster-based routing in LEACH and static routing in PEGASIS. Khan *et al.* [18] addressed the problem of sensor node movement in wireless body area sensor networks using a dynamic routing algorithm. Owing to diverse activities of humans, the positions of sensor nodes on the human body change every second. Therefore, packet and energy losses occur during transmission when nodes use the static routing algorithm. The authors solved this problem using the information of the residual energies of nodes, hop count to sink distance and throughput when nodes select the next hop node to forward data. Baker *et al.* [19] applied the GreeDi routing protocol to the ad hoc on-demand distance vector (AODV) in vehicular ad hoc networks (VANET), named GreeAODV, to achieve an energy-efficient routing protocol in the next hop selection. They modeled city map-based VANET scenarios and demonstrated that the proposed algorithm was better than the original AODV. Oubbati *et al.* [20] proposed an energy-efficient routing protocol, named energy connectivity-aware data delivery, in the flying ad hoc network. They ensured the connectivity of the proposed routing protocol by using the information on unmanned aerial vehicles (UAVs), such as their speed and location, to minimize the packet loss caused by the movement of UAVs.

There are some studies on maximizing data aggregation and network lifetime. Oubbati *et al.* [21] addressed the trade-off between efficient data aggregation and total link cost minimization. They used a comprehensive weight, named weighted data aggregation routing strategy, for solving the trade-off. By overlapping the paths of the nodes in a cluster-based WSN, they maximized the efficiency of data aggregation and prolonged the network lifetime. Ardakani *et al.* [22] presented a data-aggregation-aware efficient-routing algorithm in which the mobile agent received data from sensor nodes and aggregated and transmitted the data to the sink. They solved the delay and packet loss in routing protocols using the movement scheme of the mobile agents. Haseeb *et al.* [23] addressed the security issues in applying the conventional routing algorithm to a large-area Internet of things. They proposed light-weight structure-based data aggregation routing, which is a secure protocol that uses in-route data aggregation for routing data in the conventional routing protocols. Yazici *et al.* [24] presented a fusion-based framework to reduce the amount of data to be transmitted over the wireless multimedia sensor network by intra-node processing. They designed a sensor node to detect objects using machine learning techniques and proposed a method for increasing the accuracy while reducing the data amount. For sensor network routing, a new cluster-based routing algorithm that consume less power was presented. Clustering is one of the important techniques for topology control, effective data aggregation and energy-efficient routing in WSN.

Many researchers have applied machine learning techniques to obtain the optimal routing path with low overhead and cost. Chang *et al.* [25] applied the k-means algorithm and a genetic algorithm for multi-objective optimization. The sensor nodes in the network were clustered using the k-means algorithm. They constructed a fitness function of the genetic algorithm to maximize the network lifetime. Thangaramya *et al.* [26] presented a neuro-fuzzy-based energy-efficient clustering algorithm. In neuro-fuzzy, they used a membership function comprising the communication distance and energy information of nodes to use the energy-efficient clusters to minimize packet loss. Guo *et al.* [27] proposed an energy-efficient routing protocol based on a reinforcement learning algorithm. The nodes were reinforced to calculate the optimal routing path using a reward policy to maximize the energy efficiency and lifetime of the network. Wang *et al.* [28] used the ant colony optimization (ACO) algorithm to address the mobile sink wireless sensor network routing protocol. They proposed an improved ACO algorithm that considered not only the time and energy but also the distance between the selected cluster head (CH) and a mobile sink to calculate the optimum mobility trajectory.

El Alami and Najid [29] proposed the LEACH-based fuzzy cluster head selection algorithm. They determine the chance value using the membership function that consists of residual energy, expected efficiency and the closeness to base station. The nodes which have the higher chance value are selected as CHs in a round. Lee and Teng [30] improve the LEACH algorithm using fuzzy logic in mobile sensor network. The change of location of the nodes in network causes packet losses so they use the membership function that is made of residual energy, the movement speed and pause time of nodes. By the membership function, the chance values of all nodes to elect the CH nodes are calculated. El Alami and Najid [31] proposed an enhanced clustering hierarchy (ECH) approach to achieve energy efficiency in WSNs by using sleeping-waking mechanism for overlapping and neighboring nodes. Thus, the data redundancy is minimized and then network lifetime is maximized. Sert and Yazıcı [32] proposed the modified clonal selection algorithm (CLONALG-M) applied to determine the approximate form of the output membership functions to improve the performance of rule-based fuzzy routing. Fuzzy approach is superior to well-defined methodologies, especially where boundaries between clusters are unclear. They derived the optimal solution by using the initial membership function and iterative experiment.

Some studies have focused on data aggregation-based energy efficient routing in WSNs. Sensing data routing in network aggregation provides a better solution in terms of the reduced number of messages, high aggregation rate and reliable transmission. Zhang *et al.* [33] proposed the data aggregation mechanism supported by dynamic routing. Nodes in network select the neighbor node as next hop, which has the minimum value of function that is made of residual energy, hop count and the size of remained buffer. Li *et al.* [13] presented differentiated data aggregation routing (DDAR) that makes different QoS (Quality of Service) routes to sink node based on aggregation threshold and aggregation deadline. Most of conventional data aggregation-based routing algorithms are generally utilizing tree structure or hierarchical clustering architecture to aggregate the data and to find out the optimum route to the sink. However, they have not considered network-wise data aggregation possibilities and corresponding energy consumption for different sensor types, in which they depend on type-dependent neighbor

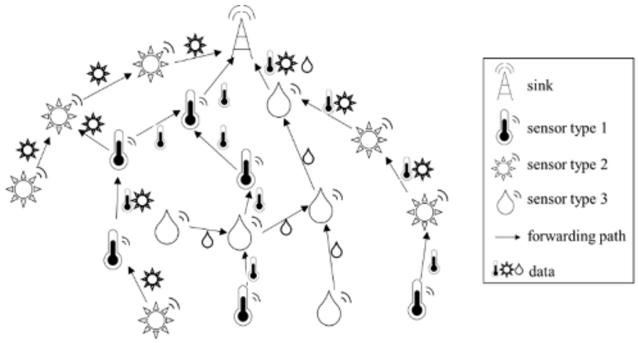


FIGURE 1. WSN model with multiple sensor types.

relationship and aggregation degrees of paths. To capture network-wise dynamics, machine learning based adaptive routing path evaluation mechanism is required. In this article, we propose a Q-learning-based routing algorithm to obtain the best next-hop node to maximize the efficiency of in-network processing. In addition, the network-wise energy consumption for communication and the residual energy of every intermediate node are also considered.

III. PROPOSED MODEL

A. NETWORK MODEL

In this study, we assume that various types of sensors, such as temperature sensors, humidity sensors and photosensors, are deployed in a field, as depicted in Fig. 1. Each sensor type has different sensing intervals based on various operating requirements. A sensor node stores its observed data and any received data from its one-hop neighbor nodes in its buffers. Each node maintains multiple sensor-type-dependent buffers. The same-sensor-type data among neighbor nodes have strong correlation. Therefore, the data of the same-type sensors can be aggregated at each node before being forwarded, as depicted in Fig. 1 [34]. Each sensor node periodically forwards its stored data to one of its one-hop neighbor nodes based on the proposed reinforcement-learning-based routing algorithm; eventually, the data are delivered to the sink node.

A sink node periodically broadcasts a Hello packet with an incremental sequence number and an initial zero hop count value. As in the publish/subscribe model in the WSN [35], a sink node declares its interest in the Hello packet. When a sensor node receives a Hello packet, it increases the hop count by 1 and rebroadcasts it to its neighbors. When a sensor node receives a Hello packet that has the same sequence number but a larger hop count, it simply discards the packet. With operation, all sensor nodes in the network always know the minimum hop count to the sink node. The proposed Q-DAEER is designed to apply to the flat network as in Fig. 1. However, the concept of Q-DAEER can be extended to the cluster-based hierarchical network architecture for inter-cluster routing between cluster heads.

B. FUNCTIONAL MODEL

A schematic of the proposed method is depicted in Fig. 2. To reduce the energy consumption for environment sensing,

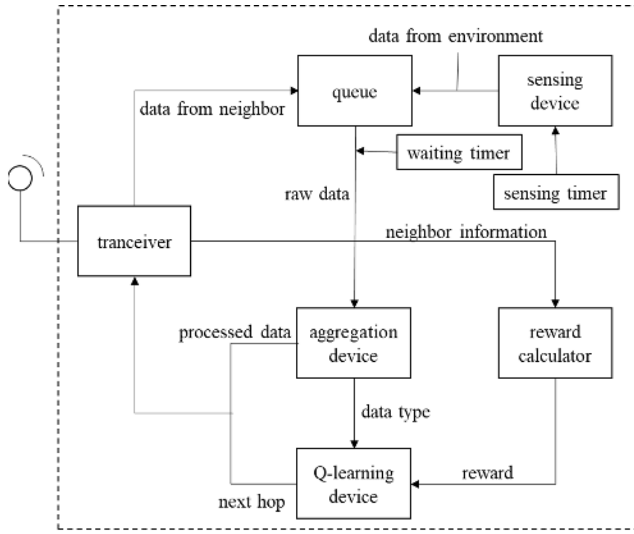


FIGURE 2. Schematic of the proposed system.

sensors periodically sense the environment based on a predefined sensing schedule for each sensor type. When the sensing timer expires, the sensing module collects the data from the environment and saves them in its sensor-type queue. Each node can receive any sensor-type data from its neighbor nodes through a transceiver and stores the data in the queue for the corresponding sensor type. Data collection at each node can be performed during a predefined waiting time for each sensor type. Depending on the latency requirement for each sensor type, the waiting time at the queue can be determined. When the waiting timer expires, the stored data in the queue are passed to the aggregation module. In the aggregation module, all raw data of each sensor type measured by the node itself and collected from neighbor nodes are aggregated using the aggregation model described in Section III.D. The aggregated data for each sensor type are forwarded to the best neighbor node, which is determined using the proposed Q-learning algorithm (see Section IV). After the neighbor node receives the data, it responds with the ACK (acknowledgement) packets, which have the status information of the data aggregation degree, hop count to the sink node, energy-related values and the location of a node. Based on the response, the sending node calculates the reward to update the Q-table for the corresponding sensor type.

C. SENSING AND DATA TRANSMISSION MODEL

In this section, we introduce the WSN sensing and data transmission model of the proposed system. In WSN, the sensor node is composed of a sensor part for monitoring the surrounding environment and a transceiver part for transmitting and receiving data. It is assumed that each sensor node does not continuously sense the surrounding environment, and the required sensing time and sensing interval for each sensor type are predetermined. The sensing start time at each node does not need to be synchronized with other nodes so that

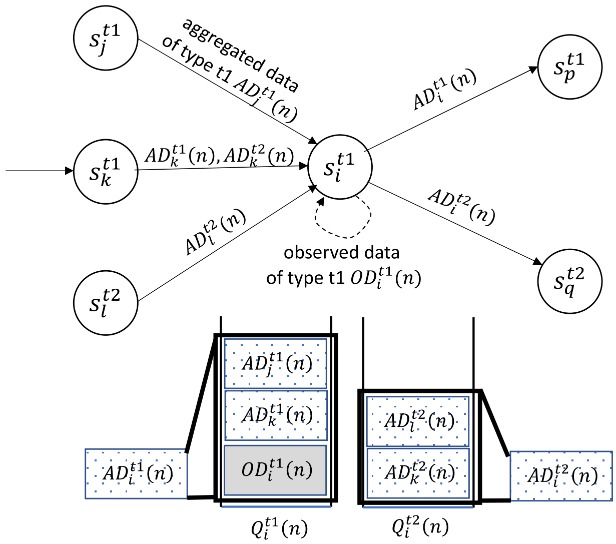


FIGURE 3. Data aggregation and transmission system model.

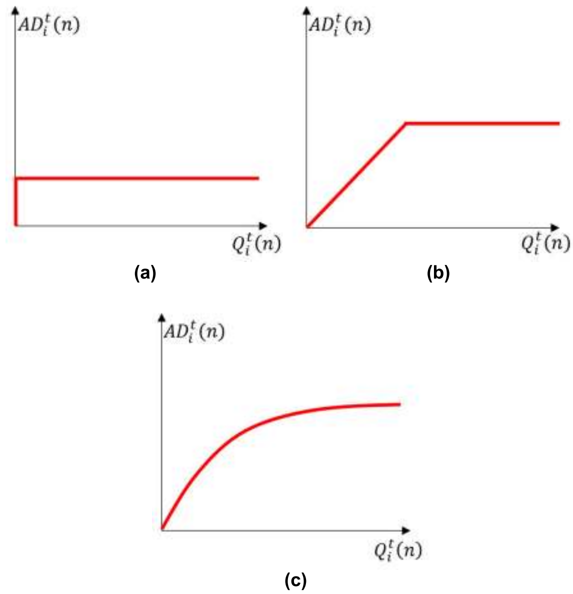


FIGURE 4. Data aggregation models (a) Representative aggregation (b) Lossy compressive aggregation (c) Lossless aggregation.

asynchronous sensing method is used. On the other hand, WSN transceivers generally use multi-mode (e.g., active, idle and sleep) operation for energy-efficiency, in which there exists the transceiver wakeup time synchronization issue with neighbor nodes. In the synchronous transceiver wakeup method, complex clock synchronization implementation and high control packet overhead exist. In the asynchronous method, there is high overhead for obtaining the wakeup schedules of neighboring nodes in advance and packet delivery latency can be higher than that of the synchronous method.

In Fig. 3, it is assumed that each sensor node is equipped with one sensor type. Notation s_i^t represents sensor node i with sensor type t. A sensor node can have multiple types of sensors, as $s_i^{t1,\cdots,tk}$. There are K different sensor types in the WSN, and each node has K queues to separately store data for various sensor types. Note that even if the sensor node has only one sensor, it should have K queues because it can be used as a relay node for any type of data. Fig. 3 shows the process of performing data aggregation on the routing path to the sink node. It was assumed that s_i^{t1} node is determined as the next node on the path to the sink node by the previous nodes. As depicted in Fig. 3, at the nth time step, sensor node s_i^{t1} measures the environment and has the observed data of sensor type $t1$, $OD_i^{t1}(n)$. It also receives aggregated data for each sensor type from its neighbor nodes. $AD_j^{t1}(n)$ indicates the aggregated data of type $t1$ at time step n from neighbor node j. During time step n, node s_i^{t1} stores all data (the received aggregated data and its local observed data) in sensor type queues $Q_i^t(n)$, $t = t1, \cdots, tK$. At the end of time step n, the node aggregates the stored data as $AD_i^t(n)$, $t = t1, \cdots, tK$, and then it forwards the aggregated data of each type to the selected neighbor nodes.

Fig. 4 illustrates the sensing and transmission of data in the proposed system model. Generally, to save energy, instead of continuous sensing, sensor nodes in the WSN sense the environment at a predefined sensing interval. In our model, we defined the sensing time and sensing interval for each sensor type t as ST^t and SI^t, respectively. For data aggregation for in-network processing, each node must wait for a certain amount of time to possibly receive the same type of data from the neighbor nodes. A longer waiting time for data aggregation results in larger latency for data delivery to the sink node. Because the level of time delay required for each sensor-type data may be different, the waiting time is set differently for each type in this model. WT^t represents the waiting time for sensor type t data aggregation. Typically, WT^t is larger than ST^t and, during sensing interval SI^t, we have multiple WT^t time steps. All nodes need not be time synchronized; they can start their schedules independently at any time. As depicted in Fig. 4, at the nth waiting time step, if there is a scheduled sensing time, the s_i^t node measures the environment during ST^t and obtains data $OD_i^t(n)$. The node will wait until the waiting timer expires to receive aggregated data from its neighbors. In Fig. 4, s_i^t receives $AD_a^t(n)$ and $AD_b^t(n)$ from nodes a and b, respectively. At the end of $WT^t(n)$, s_i^t aggregates all stored data of its type t queue $Q_i^t(n)$ and sends them to the next neighbor. When s_i^t receives aggregated data from the neighbor before the next sensing time, the node will wait for aggregated data from neighbors until the waiting timer expires.

The queue state and aggregated data size of the s_i^t sensor node at time step n are computed as follows:

$$Q_i^t(n) = OD_i^t(n) + \sum_{j \in N_i} AD_j^t(n) \tag{1}$$

$$AD_i^t(n) = DA\left\{Q_i^t(n)\right\} \tag{2}$$

where N_i is the set of neighbor nodes of node i, and $DA\{\ \}$ is the data aggregation function (explained in Section III.D).

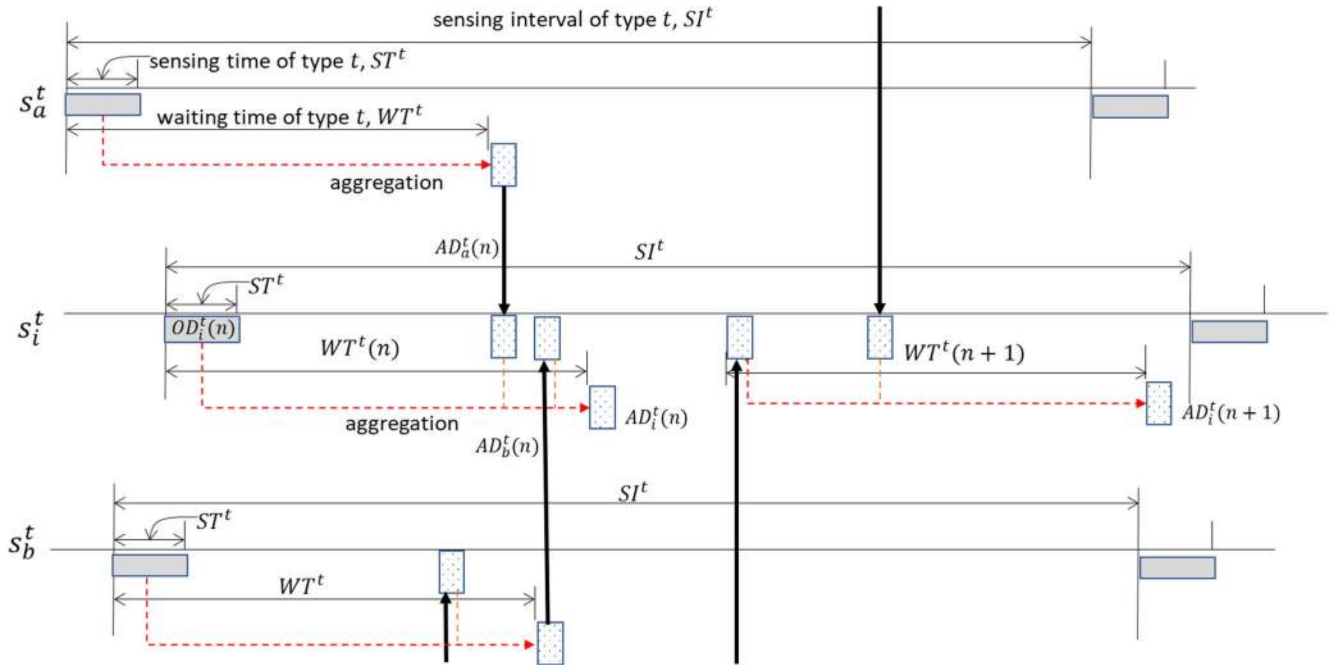


FIGURE 5. Schematic of sensing and data forwarding procedures (type t data only).

In Eq. (1), if there is no scheduled sensing time for type t at time step n, then $OD_i^t(n) = 0$.

The required energy for data transmission is generally proportional to the size of the aggregated data and the distance between the sender and receiver if the sensor nodes can control the transmission power. The required reception energy depends on the size and decoding of the data. The required energy for data aggregation is proportional to the queue state [36].

The total transmission energy required by node i at the nth time step is

$$E_i^{TX}(n) = \sum\nolimits_{\forall t} \frac{AD_i^t(n)}{B} \left\{ P_{txElec} + P_{amp} \left(\frac{d_{i-n_*^t}}{d_{max}} \right)^{\beta} \right\} \quad (3)$$

where B is the nominal bit rate; P_{txElec} is the transmission power; P_{amp} is the amplifier power; d_{max} is the maximum distance for communication at each node;$d_{i-n_*^t}$ is the distance between node i and the selected next neighbor node for type t using the proposed routing algorithm, and β is the path loss exponent ($\beta = 2$ for free space).

The total reception energy required by node i at the nth time step is

$$E_i^{RX}(n) = \sum\nolimits_{\forall t} \left\{ \frac{AD_i^t(n)}{B} P_{rxElec} + AD_i^t(n) E_{decBit} \right\} \quad (4)$$

where P_{rxElec} is the reception power, and E_{decBit} is the decoding energy per bit.

The total energy required for data aggregation by node i at the nth time step is

$$E_i^{DA}(n) = \sum\nolimits_{\forall t} Q_i^t(n) E_{aggBit} \quad (5)$$

where E_{aggBit} is data aggregation energy per bit.

D. DATA AGGREGATION MODEL

Owing to the high node density in sensor networks, similar data are sensed by many nodes, which results in redundancy in the sense data. Using data aggregation techniques, temporal and spatial redundancies can be reduced while routing packets from the source to the sink [37]–[39].

In this study, we consider three different types of data aggregation models. The first is a representative aggregation model, in which the sink node represents only a representative value. The typical mathematical functions are sum, average, maximum, minimum or median. In this model, regardless of the cumulative queue state size, the aggregated data can have a unit packet size, as depicted in Fig. 5(a). The second model is the lossy compressive aggregation model. In this model, the sensed data from multiple sensors can be represented by the limited size of the feature vector, in which various types of dimension reduction techniques with information loss can be applied. As depicted in Fig. 5(b), when the queue state is less than the feature vector size of the transformed domain, the data in the queue are transmitted without further aggregation. The third model is lossless aggregation, in which the sink node can reconstruct the raw data from the aggregated data without any loss. In this study, we modeled this type of aggregation using a log function, as depicted in Fig. 5(c). The three different data aggregation models are represented mathematically as follows:

$$DA_{model1}\left\{Q_i^t(n)\right\} = \begin{cases} U_{m1}^t & if\ Q_i^t(n) > 0 \\ 0 & if\ Q_i^t(n) = 0 \end{cases} \quad (6)$$

TABLE 1. System model parameters.

Parameter	Symbol	Parameter	Symbol
Sensor node i with sensor type t	s_i^t	Observed data of sensor node i with sensor type t at time step n	$OD_i^t(n)$
Aggregated data by sensor node i for type t at time step n	$AD_i^t(n)$	Queue state of sensor node i for type t at time step n	$Q_i^t(n)$
Sensing time for type t	ST^t	Sensing interval for type t	SI^t
Waiting time for type t for data aggregation	WT^t	Set of neighbor nodes of node i	N_i
Total energy required for transmission by sensor node i at time step n	$E_i^{TX}(n)$	Total required energy for reception of sensor node i at time step n	$E_i^{RX}(n)$
Total energy required for data aggregation by sensor node i at time step n	$E_i^{DA}(n)$	Transmission energy coefficient	C_{TX}
Reception energy coefficient	C_{RX}	Data aggregation energy coefficient	C_{DA}
Path loss exponent	β	Distance between node i and the selected next neighbor node for type t	$d_{i-n_*^t}$
Unit packet size of type t for aggregation model m	U_m^t	Number of aggregated data packets in node i's queue at time step n	$DP_i(n)$

$$DA_{model2}\left\{Q_i^t(n)\right\} = \begin{cases} U_{m2}^t & \text{if } U_{m2}^t < Q_i^t(n) \\ Q_i^t(n) & \text{if } 0 < Q_i^t(n) < U_{m2}^t \\ 0 \text{ if } Q_i^t(n) = 0 \end{cases} \tag{7}$$

$$DA_{model3}\left\{Q_i^t(n)\right\} = \begin{cases} U_{m3}^t \times \log_2(DP_i(n)+1) & \text{if } 0 < DP_i(n) \\ 0 & \text{if } DP_i(n) = 0 \end{cases} \tag{8}$$

where U_{m1}^t, U_{m2}^t and U_{m3}^t are the unit packet sizes for the first, second and third models, respectively; $DP_i(n)$ is the number of aggregated data packets in the queue of node i.

The data aggregation model is designed based on the WSN application objectives and sensor data types. It means that the actual shapes of models can be different depending on the real applications and used aggregation methods. Table 1 lists the system model parameters defined in this study.

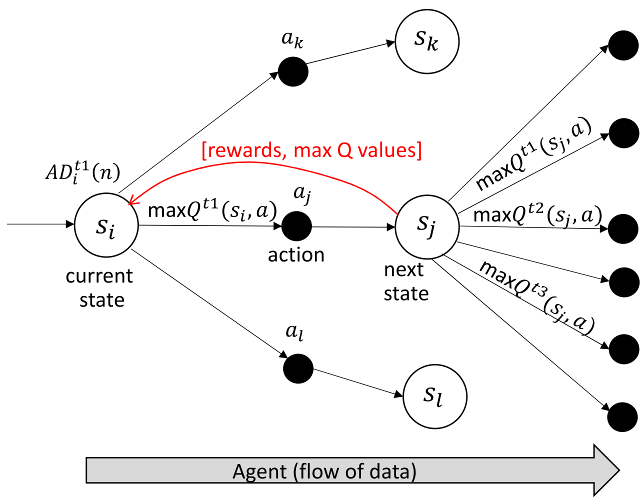


FIGURE 6. Q-learning model for the proposed system.

IV. Q-LEARNING-BASED DATA-AGGREGATION-AWARE ENERGY-EFFICIENT ROUTING PROTOCOL

Reinforcement learning methods are essential to solve optimal control problems using on-line measurements by interacting with an environment. The objective of RL is to maximize the reward of an agent by taking a series of actions in response to a dynamic environment. RL can be applied to the WSN routing problem because it can capture the dynamics of the network and environment conditions efficiently, in which the action at each sensor node is the selection of the next node for forwarding the sensing data to the sink node. Q-learning is a model-free value-based RL algorithm that is used to obtain the optimal action-selection policy using a Q value function. The Q value (quality value) represents how useful a given action is in gaining some future reward. Q-learning uses temporal differences (TD) to estimate the expected Q value through episodes with no prior knowledge of the environment. Q-learning is defined using an agent, a set of states **S** and a set of actions **A**. By performing an action $a \in \mathbf{A}$, the agent transitions from one state to another. The agent in state s interacts with the environment with action a to learn the environment, while depending on the outcome, to acquire reward r. The decision goal for selecting one of the actions in the given state is to maximize the expected sum of weighted rewards, which include the current immediate reward and future discounted rewards [40].

In the proposed Q-learning system for WSN routing, the agent is considered as a network-wide data flow. In the conventional single-agent approach, a centralized network controller acts as an agent that can observe the global conditions of the entire network and control the packet transmission at each sensor node. This central agent approach requires a large overhead and makes it difficult to know the status of the entire network in real time. In the proposed system, there is no explicit central agent; instead, cooperative information exchange among neighbor nodes ensures that each node knows the network-wide state transition behaviors. As shown in Fig. 6, the flow of data in the WSN is an agent, and each sensor node represents a state. When the type t waiting timer of sensor node i expires, it must select the next neighbor node to forward the aggregated data of type t. In this case, the current state is s_i; the actions at the current state are the list of neighbor nodes; the next state will be node s_j, to which the aggregated data of type t are forwarded. The states and actions are defined as follows:

$$\begin{aligned} &\mathbf{S} = \{s_1, s_2, \cdots, s_N\} \\ &\mathbf{A} = \{A_1, A_2, \cdots, A_N\}, \quad A_i = \left\{a_j = s_j | s_j \in N_{s_i}\right\} \end{aligned} \tag{9}$$

where N is the number of sensor nodes and N_{s_i} is the set of neighbor nodes of node s_i.

In Q-learning, the Q-table helps in finding the best action for each state, in which the action value function $Q(s, a)$ returns the expected sum of the current and future rewards when action a is performed at state s. This function can be estimated through iterative update using the Bellman equation.

Suppose that the agent selects action a in state s, observes reward R and enters new state s'. Then the action value function (Q-value), $Q(s, a)$, is updated as follows:

$$Q(s, a) = (1-\alpha) Q(s, a) + \alpha \left\{R + \gamma \cdot Q\left(s', a\right)\right\} \tag{10}$$

where α is the learning rate and γ is the discount factor for the future reward.

To achieve balance between exploitation and exploration, the epsilon-greedy strategy is generally used to select action a^* in state s, as in Eq. (11). The epsilon-greedy strategy, in which epsilon refers to the probability of choosing to explore, exploits most of the time with a small chance of exploring:

$$a_*|s = \begin{cases} \underset{a}{\operatorname{argmax}}\, Q(s, a) & \text{with probablity } 1-\epsilon \\ \text{any action } a & \text{with probability } \epsilon \end{cases} \tag{11}$$

In Q-DAEER, we perform data-type-dependent action selection and Q-table updating. Fig. 6 depicts a Q-learning scenario for WSN routing. In state s_i (sensor node i), suppose the waiting timer for type $t1$ expires so that the data in $Q_i^{t1}(n)$ aggregate into $AD_i^{t1}(n)$. In Fig. 6, the agent takes the best action that has the maximum action value for type $t1$ of the current Q-table. The best action for the given state can be different for each data type t.

The action value of action a in state s is represented as a vector, as in Eq. (12), to capture sensor-data-type-dependent expected rewards for each action:

$$\boldsymbol{Q}(s, a) = \begin{bmatrix} Q^{t1}(s, a) \\ Q^{t2}(s, a) \\ \vdots \\ Q^{tK}(s, a) \end{bmatrix} \tag{12}$$

where K is the number of sensor types. The best action for type t data forwarding in the given state s is defined as follows:

$$a_*^t | s = \underset{a}{\operatorname{argmax}}\, Q^t(s, a) \tag{13}$$

As depicted in Fig. 6, after taking the action (forwarding the aggregated data of type t) in the current state s, the agent state changes to the new state s'(the receiving sensor node of the forwarded data); the rewards are given to the current state s; the Q-table of the action taken for state s is updated. Because our Q-learning process is not controlled centrally and is performed in a distributed manner at each sensor node, the current state node s does not have the Q-table of the next state to update its Q-table using Eq. (10). In addition, state s does not explicitly know the reward for the action taken. In the data-aggregation-aware energy-efficient routing, reward R for the action in Eq. (10) represents the effectiveness of data aggregation and energy efficiency at the next node selection, and it is computed at the next state (next node). Therefore, in this study, when the next node responds the receipt of the aggregated data to the sender it also includes its maximum Q-values and the computed reward $\boldsymbol{R}$.

Because the agent acts based on the Q-value updated after the reward, it is essential to set the reward policy to determine an optimum solution for the Q-learning algorithm. We define reward $\boldsymbol{R}$ for the proposed routing algorithm as a function of rewards for the data aggregation degree, node energy status and hop count to the sink node. The data aggregation reward for type t, r_{DA}^t, is defined as in Eq. (14), and it is computed by the next node s' after it sends the received $AD_s^t(n)$ data to its queue $Q_{s'}^t(n)$ and aggregates the queued data of type t into $AD_{s'}^t(n)$.

$$r_{DA}^t = \begin{cases} \dfrac{Q_{s'}^t(n)}{AD_{s'}^t(n)} - 1 & \mathit{if}\ \dfrac{Q_{s'}^t(n)}{AD_{s'}^t(n)} - 1 < r_{DA}^{max} \\ r_{DA}^{max} & \mathit{else}\ \dfrac{Q_{s'}^t(n)}{AD_{s'}^t(n)} - 1 \geq r_{DA}^{max} \end{cases} \tag{14}$$

where r_{DA}^{max} is the maximum reward for data aggregation. In s', when the data aggregation degree (ratio between the raw and aggregated data sizes) for type t is high, reward r_{DA}^t is also high. The data aggregation reward is type dependent. When node s forwards the type t data, r_{DA}^t can be computed directly. However, the aggregation rewards for other t' types cannot be computed directly because node s did not send other types of data at this time step. In this study, we estimate the expected rewards for other types. The estimation of the expected reward for other t' types, $\hat{r}_{DA}^{t'}$, is simply defined as the most recent $r_{DA}^{t'}$ at node s'. The data aggregation reward vector ($\boldsymbol{R_{DA}}$) for all data types is defined using (15).

$$\boldsymbol{R_{DA}} = \begin{bmatrix} r_{DA}^t = \dfrac{Q_{s'}^t(n)}{AD_{s'}^t(n)} \\ \vdots \\ \hat{r}_{DA}^{t'} = r_{DA}^{t'}\left(n^-\right) = \dfrac{Q_{s'}^{t'}\left(n^-\right)}{AD_{s'}^{t'}\left(n^-\right)} \end{bmatrix} \tag{15}$$

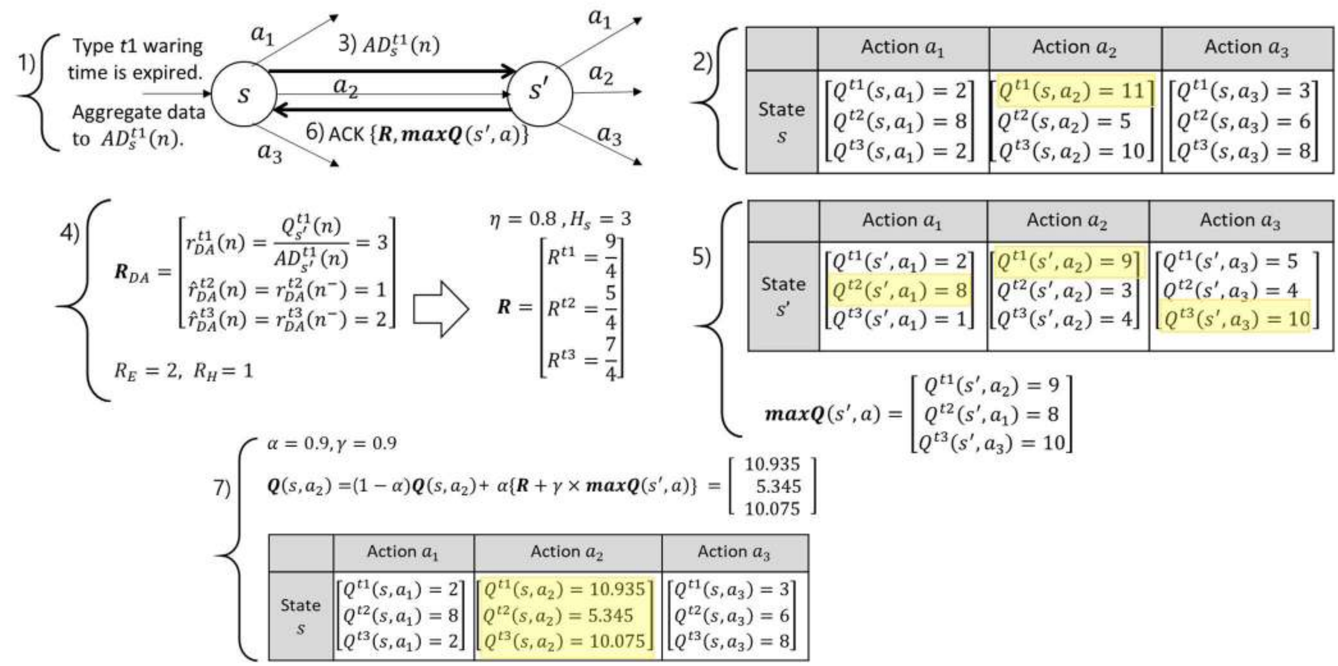


FIGURE 7. Example scenario for the proposed Q-DAEER learning process.

where t is the data-type node s sent and n^- is the most recent time step at which node s' computed $r_{DA}^{t'}$.

We have defined the type-independent energy status reward. The energy status reward (R_E) is defined as follows:

$$R_E = \frac{E_{s'}^r(n)}{E_{s'}^r(0)} - \left(\frac{d_{s-s'}}{d_{max}}\right)^\beta \quad (16) \tag{16}$$

where $E_{s'}^r(n)$ and $E_{s'}^r(0)$ are the residual energies of the next node s' at the nth and 0th time steps, respectively; $d_{s-s'}$ is the estimated distance between nodes s and s' (estimated at node s' using any distance estimation techniques);d_{max} is the maximum transmission range of the sensor nodes; and β is the path loss exponent (in free space $\beta = 2$). When the remaining energy of the next state node is relatively large and the distance between the next and current state nodes is short (which means that the energy requirement for transmission is low), the action selection is efficient in terms of energy. Consequently, the energy state reward increases. This reward policy can reduce the energy consumption of the entire network and increase the network lifetime by evenly distributing the energy consumption at each node.

To forward data to the sink, the reward should be smaller than the maximum Q-value of the parent hop count node. However, the fixed reward for all nodes in the network has a higher probability of backwarding the nodes that are away from the sink. An additional discount factor for the reward of the nodes is necessary to prevent backwarding. Reward **R** for action a in state s is finally computed as follows:

$$\boldsymbol{R} = \begin{cases} \eta^{H_s} \times (\boldsymbol{R_{DA}} + R_E \times \vec{1}) & \textit{if } s' \textit{is not a sink} \\ R_s \times \vec{1} & \textit{else} \end{cases} \tag{17}$$

where H_s is the hop count of node s, $\vec{1}$ is the K-dimensional vector with all 1s, R_s is the sink node reward and η is the discount factor for the reward in range [0, 1].

When node s receives reward $\boldsymbol{R}$, it needs to update its Q-table. To update its action value function $Q(s, a)$, it requires $Q(s', a)$ of the next state node. As explained previously, in our proposed mechanism, when the next node s' receives an aggregated data packet, it replies with the ACK packet, in which the reward vector $\boldsymbol{R}$ of Eq. (17) and $\mathbf{max}\,\boldsymbol{Q}(s', a)$ vector are included. Therefore, node s can update its Q-table based on the ACK packet information. The $\mathbf{max}\,\boldsymbol{Q}(s', a)$ vector includes the maximum Q-value for each data type at the next node s' as follows:

$$\mathbf{max}\,\boldsymbol{Q}(s', a) = \begin{bmatrix} \max\limits_{\forall a} Q^{t1}(s', a) \\ \max\limits_{\forall a} Q^{t2}(s', a) \\ \vdots \\ \max\limits_{\forall a} Q^{tK}(s', a) \end{bmatrix} \tag{18}$$

The general Q-table update rule of Eq. (10) can be represented in vector form as follows:

$$\boldsymbol{Q}(s, a) = (1-\alpha)\boldsymbol{Q}(s, a) + \alpha\left\{\boldsymbol{R} + \gamma + \mathbf{max}\,\boldsymbol{Q}(s', a)\right\} \tag{19}$$

Fig. 7 illustrates a scenario for the proposed Q-DAEER learning procedure.

1) At node s, the waiting timer of type $t1$ expires at time step n, and then node s aggregates data in queue $Q_s^{t1}(n)$ to $AD_s^{t1}(n)$.
2) Node s selects action a_2 (node s') that has the maximum Q-value for type $t1$ of state s Q-table.

3) Based on action a_2, node s forwards the aggregated data to node s'.
4) Node s' calculates reward vector $\boldsymbol{R}$.
5) Node s' derives $\max \boldsymbol{Q}\left(s', a\right)$ vector from its Q-table in the form of Eq. (18).
6) Node s' replies to ACK including $\left\{\boldsymbol{R}, \max \boldsymbol{Q}\left(s', a\right)\right\}$.
7) Node s updates $\boldsymbol{Q}\,(s, a)$ vector using Eq. (19).

Table 2 shows the complexity and overhead analysis of the proposed algorithm compared with two other WSN routing methods. The first compared algorithm is the shortest path routing using the proposed data aggregation model at each node on the path. The second one is the shortest path routing without data aggregation. The analysis has been conducted in terms of complexity, queue management overhead, control message overhead and time delay.

V. SIMULATION RESULTS AND PERFORMANCE EVALUATION

In this section, we evaluate and analyze the performance of the proposed Q-DAEER routing protocol in terms of its energy consumption, network lifetime, average hop count and decrease in data size. We implemented the simulation environments using MATLAB R2019a to compare the proposed routing algorithm with the conventional routing algorithms. The simulation parameters and values used in this study are listed in Table 3. We used the random-type grid topology for the WSN, in which sensor nodes were deployed in the form of a grid, as depicted in Fig. 8 (an example topology), and each sensor node had only a single-type sensor module that is randomly selected. The characteristics of the three types of sensor modules are summarized in Table 4. 77 sensor nodes were deployed in the sensing area. The initial energy level of nodes followed a uniform distribution with [2J, 2.5J]. The maximum transmission range of nodes was assumed to be 150 distance units (du). The unit packet sizes for data aggregation model-1, -2 and -3 given by Eqs. (6)–(8) were proportional to the observed data size by each sensor type. The transmission, amplification and reception powers were 200 mW, 500 mW and 200 mW, respectively. The nominal bit rate for nodes was 6 Mbps and the energy consumptions for decoding and data aggregation were 40 nJ and 20 nJ per bit, respectively. The observed packet sizes, sensing intervals and waiting timers of all sensor types are listed in Table 4.

We implemented two conventional energy-aware WSN routing algorithms shown in Table 2 for performance comparison. In the shortest path routing (SPR) without data aggregation, to minimize energy consumption, a sensor node in the network selects the next neighbor node that has a least hop count to the sink. This results in a minimum distance between the source node and the sink node. In the shortest path routing with data aggregation (SPRwDA), when a sensor node receives the aggregated data from other nodes, the node waits until the waiting timer expires to minimize the transmission overhead. Then it aggregates all received and locally observed data together using the proposed aggregation procedure, and

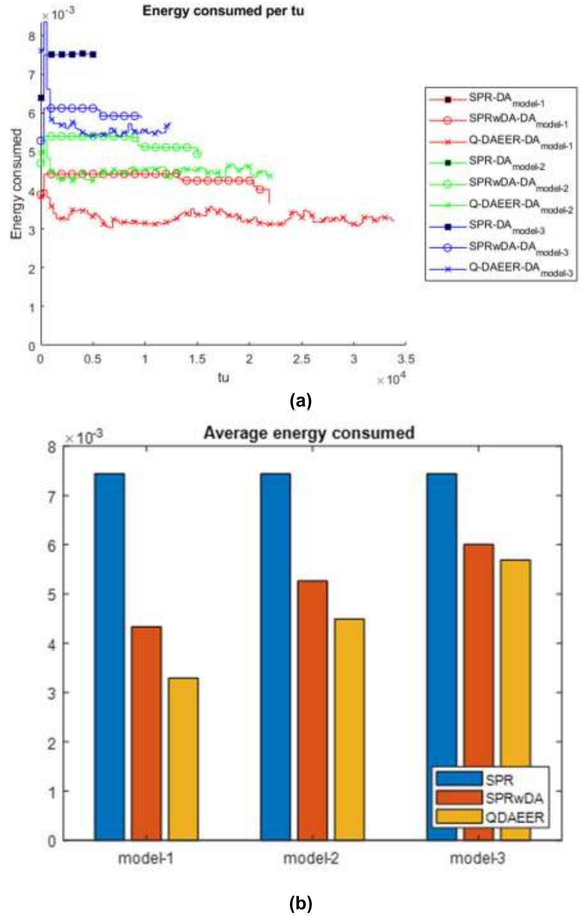


FIGURE 8. Energy consumption for nodes (a) Energy consumed per time unit (tu) (b) Average energy consumed.

it forwards the aggregated data to the next node using the shortest path routing.

We performed the simulation until half of the nodes of the one-hop neighbors of the sink were dead or some nodes in the network were isolated so that they could not transmit data to the sink. We compared the performances in terms of network-level energy consumption, number of dead nodes, network lifetime, average hop count and decrease in data size. Network-level energy consumption is the sum of energies consumed by all the sensor nodes. The number of dead nodes represents the number of sensor nodes with depleted energies. Network lifetime indicates the elapsed time until half of the nodes of the one-hop neighbors of the sink are dead or some nodes in the network are isolated so that they cannot transmit data to the sink. Average hop count is the average of the hop counts required to reach the sink node, which also approximately represents the delay from the data source to the sink node. Decrease in data size represents the amount of the reduced data size owing to data aggregation through the routing path. It represents the efficiency of data aggregation of a routing algorithm. In the simulation study, model-1, model-2 and model-3 represent the data aggregation models given by Eqs (6), (7) and (8), respectively.

TABLE 2. Complexity and overhead analysis.

Algorithm	Proposed	Shortest path with data aggregation`	Shortest path without data aggregation
Data aggregation complexity	Type-dependent data aggregation at each node is required. Depending on the aggregation models and methods, overhead degrees are different. Overhead cost for aggregation models: C(M1)<C(M2)<C(M3)	Same as the proposed algorithm.	No data aggregation overhead.
Queue management	Sensor-type dependent multiple queue management is required. O(N) complexity, N: number of sensor types	Same as the proposed algorithm.	Single queue management O(1) complexity.
Control message overhead	Sink node periodically broadcasts Hello packets. After receiving a data packet, each node replies an ACK packet including reward and Q-values.	Sink node periodically broadcasts Hello packets. After receiving a data packet, each node replies a simple ACK packet.	Same as the shortest path with aggregation algorithm.
Latency	There exists additional latency due to the type-dependent waiting time for data aggregation. Data aggregation degree is considered to determine the optimum path so that the path length in terms of hop counts can be little higher than others.	There exists additional latency due to the type-dependent waiting time for data aggregation.	Latency only depends on the hop counts of the shortest path from sensors to the sink.

TABLE 3. Simulation parameters.

Parameters	Value	Parameters	Value
Number of nodes	77, 100, 400	Number of sensor types	3
Sensor-field size	1100 × 900 (du)	Sensor topology	grid
Location of sink	{1050, 350}	d_{max}	150 (du)
$E_i^r(0)$	[2J, 2.5J]	P_{tx}	$200(mW)$
P_{amp}	$500(mW)$	P_{rx}	$200(mW)$
B	$6(Mbps)$	E_{decBit}	$40(nJ)$
E_{DABit}	$20(nJ)$	$U_{m1}^t, U_{m2}^t, U_{m3}^t$	$\{OD^t, 1.5 \times OD^t, OD^t\}$
r_{DA}^{max}	2	R_S	20
α, γ	{0.8, 0.9}		

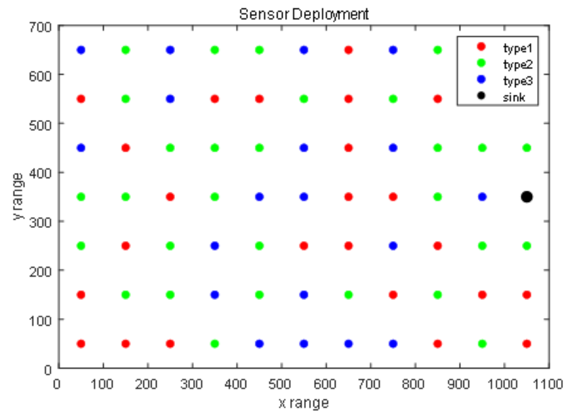


FIGURE 9. Wireless sensor network simulation environment.

TABLE 4. Sensor type dependent parameters for simulation.

Sensor type (t)	Unit packet size (OD^t)	Sensing interval (ST^t)	Waiting time (WT^t)
1	600	30	10
2	400	25	9
3	500	20	8

The comparative results of network-level energy consumption are depicted in Fig. 9. Fig. 9(a) shows the results at each time step tu (time unit). In the SPR and SPRwDA, the energy consumption at every time step is almost constant because they use the shortest routing path and it is only determined by the current network topology. Since SPRwDA uses the proposed data aggregation method before forwarding data at each node, it can be seen that the energy used is lower than that of SPR. In the proposed Q-DAEER method, the energy consumption of each sensor node in the WSN using the proposed routing algorithm is dynamic owing to the policy-based dynamic reward update rule. Initially, the energy consumption of the proposed method is higher than that of the conventional algorithms because each node needs to learn the optimal path. However, after learning, the nodes spent the least energy for all three data aggregation models. Fig. 9(b) shows the total average energy consumptions for all time steps. We can see that the proposed algorithm consumed the least energy compared with two other algorithms. In data aggregation model-1, the efficiency of data aggregation is the highest so that its average energy consumption was the lowest among all the models. For three data aggregation models, the propose Q-DAEER can reduce energy consumption by 67%∼32% compared with SPR and by 25%∼5% compared with SPRwDA.

The comparisons of the numbers of dead sensor nodes over time and the average network lifetime are shown in Fig. 10. In the case of SPR, it can be seen that the number of dead nodes increases faster than other methods due to high energy consumption. The data aggregation model-3 exhibits a faster node dead time when compared with the other two models because, as in Fig. 9(b), model-3 consumes more energy when compared with the other models. Fig. 10(b) depicts the network lifetimes when half of the nodes near the sink are dead or some of the nodes are isolated. In data aggregation model-1, the network lifetime using the proposed method is approximately 6.8∼2.5 and 1.55∼1.29 times longer than

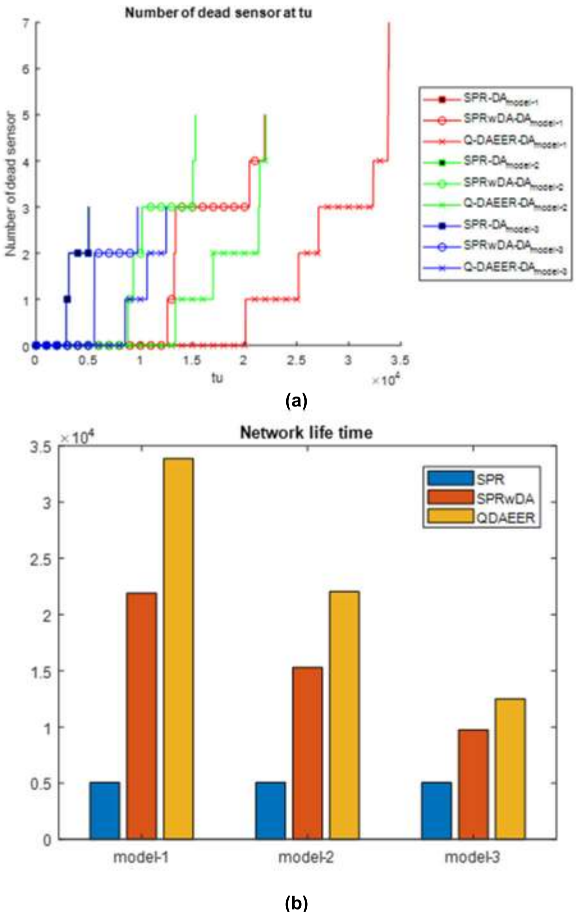


FIGURE 10. Numbers of dead sensor nodes and network lifetimes (a) Number of dead sensors per time unit (tu) (b) Network lifetimes.

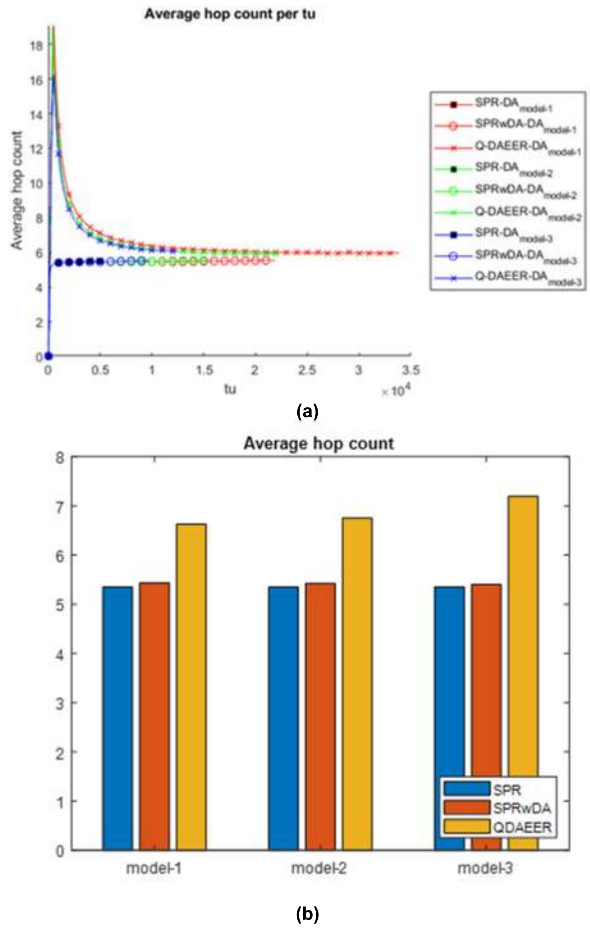


FIGURE 11. Comparison of hop count averages (a) Average hop count per time unit (tu) (b) Average hop count.

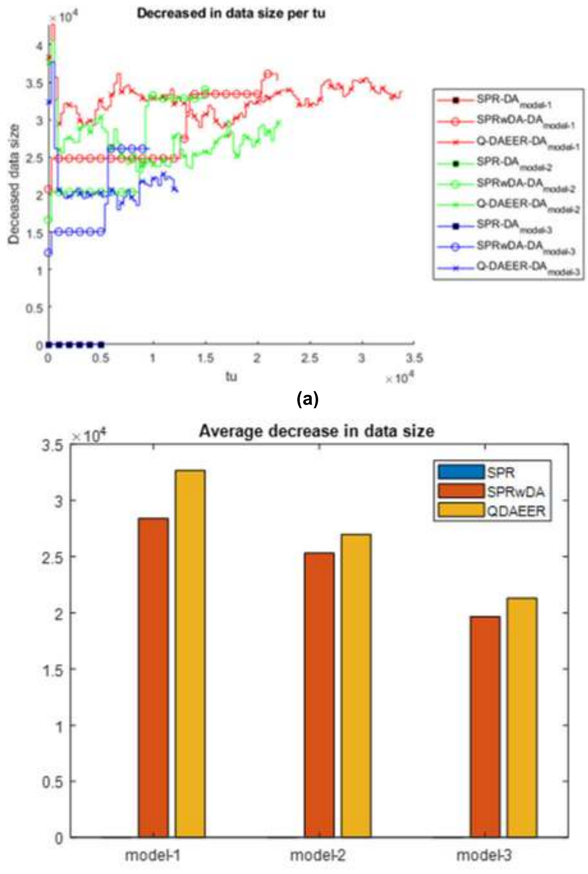
those of SPR and SPRwDA, respectively for three data aggregation models.

Fig. 11 shows the average hop count of data packets from the data source node to the sink node. The average hop count at each time unit is depicted in Fig. 11(a). In SPR and SPRwDA, because each sensor node forwards data to the neighbor that is closest to the sink node, the average hop count is almost constant and lower than that of the proposed Q-DAEER regardless data aggregation models. However, near the end of the simulation, the average hop counts of SPR and SPRwDA increase slightly because some nodes become dead owing to the depletion of their energies. In contrast, the proposed Q-DAEER method demonstrates a higher initial average hop count for reinforcement learning. In Q-learning, before the Q-table is stabilized and used, the agent needs to explore more paths. The average hop count in the proposed method decreases significantly after the initial learning period. Each sensor node dynamically learns the optimal routing path in terms of not only the hop count but also the energy consumption and data aggregation degree on the path. The Q-DAEER algorithm may choose longer paths to obtain higher expected rewards by achieving more data aggregation and energy saving. Therefore, for three data aggregation models, the average hop count of Q-DAEER is approximately 25%∼35% higher than those of SPR and SPRwDA.

A comparison of the decrease in data sizes in the network is presented in Fig. 12. Fig. 12(a) shows the decrease in the data size at each time unit. Because SPR does not perform data aggregation, the reduced data size is zero. In SPRwDA, the reduction in the data size is almost similar at each time step for roughly the first half of the network lifetime; afterward, it increases suddenly. Because SPRwDA utilizes the shortest path, the energy of some nodes close to the sink node depletes, eventually causing these nodes to stop functioning. This causes data from sensor nodes to concentrate in the remaining nodes, which can significantly reduce the data size. Therefore, the decrease in the data size increases in the second half of the simulation. However, as shown in Fig 9, this accelerates the energy shortage among the overloaded nodes and shortens the network lifetime. In the proposed Q-DAEER algorithm, the rewards that are given by the neighbor nodes consider the energy level and degree of data aggregation so that nodes always dynamically determine the best path. The results indicate that the proposed algorithm can obtain a more optimal path to improve energy and data aggregation efficiency compared with the conventional

TABLE 5. The results of grid and random topologies for 100 and 400 sensor node cases.

Average result \ Topology		Grid topology						Random topology					
		Proposed		SPRwDA		SPR		Proposed		SPRwDA		SPR	
		100	400	100	400	100	400	100	400	100	400	100	400
Data aggregation model-1	Energy consumption(10^{-3} J)	3.3	16.2	4.3	19.6	7.4	58.3	3	14.1	3.1	15.5	6.3	52.4
	Network lifetime(tu)	33870	29478	21906	18915	5049	1375	25392	32262	22939	22300	5893	1806
	Average hop count	6.63	12.15	5.44	10.34	5.35	9.35	6.03	17.93	4.52	9.07	4.49	8.42
	Decrease in data size(10^3 bits)	32.67	137.77	28.40	137.61	0	0	30.20	133.43	29.34	119.90	0	0
Data aggregation model-2	Energy consumption(10^{-3} J)	4.5	21.8	5.3	25.7	7.4	58.3	3.7	19	3.7	19.8	6.3	52.4
	Network lifetime(tu)	22034	19517	15293	12955	5049	1375	20062	21724	16214	15650	5893	1806
	Average hop count	6.75	12.75	5.42	10.34	5.35	9.35	6.05	13.13	4.55	9.05	4.49	8.42
	Decrease in data size(10^3 bits)	27.0	129.17	25.31	126.97	0	0	28.46	121.49	25.19	112.12	0	0
Data aggregation model-3	Energy consumption(10^{-3} J)	5.7	33.3	6	34.5	7.4	58.3	4.8	30	4.6	28.7	6.3	52.4
	Network lifetime(tu)	12493	6777	9740	5515	5049	1375	11851	9435	10418	8522	5893	1806
	Average hop count	7.20	14.63	5.40	10.16	5.35	9.35	7.03	13.92	4.50	9.0623	4.49	8.42
	Decrease in data size(10^3 bits)	21.29	111.36	19.66	107.59	0	0	21.44	107.38	18.79	101.66	0	0



(a)

(b)

FIGURE 12. Comparison of data size reduction due to data aggregation (a) Per time unit (tu) decrease in data size due to aggregation (b) Average decrease in data size.

method. As depicted in Fig. 12(b), the proposed algorithm achieved approximately 20%∼10% higher data reduction ratio compared with SPRwDA for three aggregation models.

We applied a random topology in addition to the grid topology in the previous experiments in the sensor deployment topology, and also verified the scalability of the proposed algorithm by increasing the number of nodes to 100 and 400. Table 5 shows the experimental results with the compared methods. As we can see, the proposed Q-DAEER method consumed less energy and achieved longer network lifetime for both of random and grid topology at even dense node conditions.

VI. CONCLUSION

In this article, we proposed a Q-learning-based data-aggregation-aware energy-efficient routing (Q-DAEER) algorithm. To calculate the best path to maximize the lifetime and minimize energy consumption of the network, we defined a reward policy that considered the energy level, distance, hop count and the degree of data aggregation at each node. For efficient data aggregation at each node with different sensor types, we presented a data aggregation and system model in which sensor-type-dependent queue management and transmission schedule control were used. The reward functions defined in this study captured the changes in the energy node, neighbor relationship and type-dependent data aggregation dynamics of each node. In the proposed Q-DAEER algorithm, we incorporated a data-type-dependent action selection and Q-table updating algorithm. To demonstrate the applicability of the proposed algorithm to various data aggregation scenarios, we defined three different data aggregation models. We compared the performance of the proposed algorithm with that of the conventional routing protocol in terms of its energy consumption, network lifetime, average hop count and degree of data aggregation. The results indicate that the proposed algorithm can obtain a more optimal path to improve energy and data aggregation efficiencies when compared with the conventional method. We demonstrated that the proposed Q-DAEER protocol can successfully reduce the overall data transmission load and extend the lifetime of the wireless sensor network.

REFERENCES

[1] Y. Jin, K. S. Kwak, and S.-J. Yoo, "A novel energy supply strategy for stable sensor data delivery in wireless sensor networks," *IEEE Syst. J.*, vol. 14, no. 3, pp. 3418–3429, Sep. 2020.

[2] S. Wang, J. Wan, D. Li, and C. Xhang, "Implementing smart factory of industrie 4.0: An outlook," *Int. J. Distrib. Sensor Netw.*, vol. 12, pp. 1–10, Jan. 2016.

[3] T.-H. Kim, C. Ramos, and S. Mohammed, "Smart city and IoT," *Future Gener. Comput. Syst.*, vol. 76, pp. 159–162, Nov. 2017.
[4] K. Cengiz and T. Dag, "Energy aware multi-hop routing protocol for WSNs," *IEEE Access*, vol. 6, pp. 2622–2633, 2018.
[5] J. Huang, Y. Hong, Z. Zhao, and Y. Yuan, "An energy-efficient multi-hop routing protocol based on grid clustering for wireless sensor networks," *Cluster Comput.*, vol. 20, no. 4, pp. 3071–3083, Jun. 2017.
[6] P. Guo, J. Cao, and X. Liu, "Lossless in-network processing in WSNs for domain-specific monitoring applications," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2130–2139, Oct. 2017.
[7] S. A. Putra, B. R. Trilaksono, A. Harsoyo, and A. I. Kistijantoro, "Multiagent system in-network processing in wireless sensor network," *Int. J. Electr. Eng. Informat.*, vol. 10, no. 1, pp. 94–107, Mar. 2018.
[8] Z. A. Khan and A. Samad, "A study of machine learning in wireless sensor network," *Int. J. Comput. Netw. Appl.*, vol. 4, no. 4, pp. 105–112, Aug. 2017.
[9] D. P. Kumar, T. Amgoth, and C. S. R. Annavarapu, "Machine learning algorithms for wireless sensor networks: A survey," *Inf. Fusion*, vol. 49, pp. 1–25, Sep. 2019.
[10] Z. Mammeri, "Reinforcement learning based routing in networks: Review and classification of approaches," *IEEE Access*, vol. 7, pp. 55916–55950, 2019.
[11] J. Wang, Y. Gao, W. Liu, A. K. Sangaiah, and H.-J. Kim, "Energy efficient routing algorithm with mobile sink support for wireless sensor networks," *Sensors*, vol. 19, pp. 1–19, Jan. 2019.
[12] S. K. Das and S. Tripathi, "Intelligent energy-aware efficient routing for MANET," *Wireless Netw.*, vol. 24, no. 4, pp. 1139–1159, May 2018.
[13] X. Li, W. Liu, M. Xie, A. Liu, M. Zhao, N. N. Xiong, M. Zhao, and W. Dai, "Differentiated data aggregation routing scheme for energy conserving and delay sensitive wireless sensor networks," *Sensors*, vol. 18, no. 7, pp. 1–29, Jul. 2018.
[14] S. Sasirekha and S. Swamynathan, "Cluster-chain mobile agent routing algorithm for efficient data aggregation in wireless sensor network," *J. Commun. Netw.*, vol. 19, no. 4, pp. 392–401, Aug. 2017.
[15] Y. Bai, Y. Mai, and N. Wang, "Performance comparison and evaluation of the proactive and reactive routing protocols for MANETs," in *Proc. Wireless Telecommun. Symp. (WTS)*, Chicago, IL, USA, Apr. 2017, pp. 1–5.
[16] R. E. Mohemed, A. I. Saleh, M. Abdelrazzak, and A. S. Samra, "Energy-efficient routing protocols for solving energy hole problem in wireless sensor networks," *Comput. Netw.*, vol. 114, pp. 51–66, Feb. 2017.
[17] A. Razaque, M. Abdulgader, C. Joshi, F. Amsaad, and M. Chauhan, "P-LEACH: Energy efficient routing protocol for wireless sensor networks," in *Proc. IEEE Long Island Syst., Appl. Technol. Conf. (LISAT)*, Apr. 2016, pp. 1–5.
[18] R. A. Khan, Q. Xin, and N. Roshan, "RK-energy efficient routing protocol for wireless body area sensor networks," *Wireless Pers. Commun.*, vol. 116, pp. 1–13, Aug. 2020.
[19] T. Baker, M. J. García-Campos, D. G. Reina, S. Toral, H. Tawfik, D. Al-Jumeily, and A. Hussain, "GreeAODV: An energy efficient routing protocol for vehicular ad hoc networks," in *Proc. Int. Conf. Intell. Comput.*, Jul. 2018, pp. 670–681.
[20] O. S. Oubbati, M. Mozaffari, N. Chaib, P. Lorenz, M. Atiquzzaman, and A. Jamalipour, "ECaD: Energy-efficient routing in flying ad hoc networks," *Int. J. Commun. Syst.*, vol. 32, no. 18, p. e4156, Dec. 2019.
[21] O. A. Mahdi, A. W. A. Wahab, M. Y. I. Idris, A. A. Znaid, Y. R. B. Al-Mayouf, and S. Khan, "WDARS: A weighted data aggregation routing strategy with minimum link cost in event-driven WSNs," *J. Sensors*, vol. 2016, pp. 1–12, Sep. 2016.
[22] S. P. Ardakani, J. Padget, and M. De Vos, "A mobile agent routing protocol for data aggregation in wireless sensor networks," *Int. J. Wireless Inf. Netw.*, vol. 24, no. 1, pp. 27–41, Mar. 2017.
[23] K. Haseeb, N. Islam, T. Saba, A. Rehman, and Z. Mehmood, "LSDAR: A light-weight structure based data aggregation routing protocol with secure Internet of Things integrated next-generation sensor networks," *Sustain. Cities Soc.*, vol. 54, pp. 1–9, Mar. 2020.
[24] A. Yazici, M. Koyuncu, S. A. Sert, and T. Yilmaz, "A fusion-based framework for wireless multimedia sensor networks in surveillance applications," *IEEE Access*, vol. 7, pp. 88418–88434, 2019.
[25] Y. Chang, X. Yuan, B. Li, D. Niyato, and N. Al-Dhahir, "Machine-learning-based parallel genetic algorithms for multi-objective optimization in ultra-reliable low-latency WSNs," *IEEE Access*, vol. 7, pp. 4913–4926, 2019.
[26] K. Thangaramya, K. Kulothungan, R. Logambigai, M. Selvi, S. Ganapathy, and A. Kannan, "Energy aware cluster and neuro-fuzzy based routing algorithm for wireless sensor networks in IoT," *Comput. Netw.*, vol. 151, pp. 211–223, Mar. 2019.
[27] W. Guo, C. Yan, and T. Lu, "Optimizing the lifetime of wireless sensor networks via reinforcement-learning-based routing," *Int. J. Distrib. Sensor Netw.*, vol. 15, no. 2, Feb. 2019, Art. no. 155014771983354.
[28] J. Wang, J. Cao, R. S. Sherratt, and J. H. Park, "An improved ant colony optimization-based approach with mobile sink for wireless sensor networks," *J. Supercomput.*, vol. 74, no. 12, pp. 6633–6645, Dec. 2018.
[29] H. El Alami and A. Najid, "Energy-efficient fuzzy logic cluster head selection in wireless sensor networks," in *Proc. Int. Conf. Inf. Technol. Org. Develop. (IT OD)*, Rabat, Morocco, Mar. 2016, pp. 1–7.
[30] J.-S. Lee and C.-L. Teng, "An enhanced hierarchical clustering approach for mobile sensor networks using fuzzy inference systems," *IEEE Internet Things J.*, vol. 4, no. 4, pp. 1095–1103, Aug. 2017.
[31] H. El Alami and A. Najid, "ECH: An enhanced clustering hierarchy approach to maximize lifetime of wireless sensor networks," *IEEE Access*, vol. 7, pp. 107142–107153, 2019.
[32] S. A. Sert and A. Yazici, "Optimizing the performance of rule-based fuzzy routing algorithms in wireless sensor networks," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, New Orleans, LA, USA, Jun. 2019, pp. 1–6.
[33] J. Zhang, Q. Wu, F. Ren, T. He, and C. Lin, "Effective data aggregation supported by dynamic routing in wireless sensor networks," in *Proc. IEEE Int. Conf. Commun.*, Cape Town, South Africa, May 2010, pp. 1–6.
[34] R. Rajagopalan and P. K. Varshney, "Data-aggregation techniques in sensor networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 8, no. 4, pp. 48–63, 4th Quart., 2006.
[35] A. Ganesh, "Publish/subscribe model in a wireless sensor network," U.S. Patent 7 590 098, Sep. 15, 2009.
[36] H. Karl and A. Willig, *Protocols and Architectures for Wireless Sensor Networks*. Hoboken, NJ, USA: Wiley, 2007.
[37] M. Dagar and S. Mahajan, "Data aggregation in wireless sensor network: A survey," *Int. J. Inf. Comput. Technol.*, vol. 3, no. 3, pp. 167–174, 2013.
[38] X. Xu, R. Ansari, A. Khokhar, and A. V. Vasilakos, "Hierarchical data aggregation using compressive sensing (HDACS) in WSNs," *ACM Trans. Sensor Netw.*, vol. 11, no. 3, pp. 1–25, May 2015.
[39] F. Marcelloni and M. Vecchio, "An efficient lossless compression algorithm for tiny nodes of monitoring wireless sensor networks," *Comput. J.*, vol. 52, no. 8, pp. 969–987, Nov. 2009.
[40] Q. Yang, S.-J. Jang, and S.-J. Yoo, "Q-learning-based fuzzy logic for multi-objective routing algorithm in flying ad hoc networks," *Wireless Pers. Commun.*, vol. 113, no. 1, pp. 115–138, Jul. 2020.

WAN-KYU YUN received the B.S. degree from the Information and Communication Engineering Department, Inha University, South Korea, where he is currently pursuing the M.S. degree with the Multimedia Network Laboratory. His research interests include wireless sensor networks, the Internet of Things, machine learning, and deep learning.

SANG-JO YOO (Member, IEEE) received the B.S. degree in electronic communication engineering from Hanyang University, Seoul, South Korea, in 1988, and the M.S. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology, in 1990 and 2000, respectively.

From 1990 to 2001, he was a Member of Technical Staff with the Korea Telecom Research and Development Group, where he was involved in communication protocol conformance testing and network design fields. From 1994 to 1995 and 2007 to 2008, he was a Guest Researcher with the National Institute Standards and Technology, USA. Since 2001, he has been with Inha University, where he is currently a Professor with the Information and Communication Engineering Department. His current research interests include cognitive radio network protocols, adhoc wireless networks, MAC and routing protocol design, wireless networks, QoS, and wireless sensor networks.