Q-learning-based Traffic-aware Parent Selection for Wireless Powered Sensor Networks

Sol-Bee Lee
Division of Software
Hallym University
Chuncheon, South Korea
thfqla3535@hallym.ac.kr

Jung-Hyok Kwon
Smart Computing Laboratory
Hallym University
Chuncheon, South Korea
jhkwon@hallym.ac.kr

Eui-Jik Kim*
Division of Software
Hallym University
Chuncheon, South Korea
ejkim32@hallym.ac.kr

2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) | 979-8-3503-4434-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICAIIC60209.2024.10463229

***Abstract*— This paper presents a Q-learning-based traffic-aware parent selection for wireless powered sensor networks (WPSNs), abbreviated QTaPS. The QTaPS employs Q-learning to independently select power and data parents, thereby improving the efficiency of energy harvesting and data transmission in WPSNs. To this end, QTaPS uses two Q-learning agents with different reward functions to select power and data parents, respectively. In order to consider the characteristics of each type of traffic, different metrics are used in each reward function. The simulation results showed that QTaPS obtained better performance than the existing scheme in terms of aggregate throughput and average end-to-end delay.**

Keywords—energy harvesting, parent selection, Q-learning, WPSN, WPT

I. Introduction

Recently, self-sustainable wireless powered sensor networks (WPSNs), which enable permanent operation of sensor nodes without battery replacement or interruption by using wireless power transfer (WPT) technology, have been extensively studied [1–4]. Unlike traditional wireless sensor networks (WSNs), WPSNs consist of multiple sensor nodes and hybrid access points (HAPs) that supply power to sensor nodes using radio frequency (RF) signals. The HAPs also collect data from the sensor nodes and then transmit it towards the root. Accordingly, two types of traffic—data traffic and power traffic—with different characteristics coexist in WPSN. Data traffic represents data packets transmitted in the network and is transmitted through multi-hop from the sensor node to the root. On the other hand, power traffic represents RF signals transmitted in a single hop from the HAP to the sensor node to charge the sensor node. In particular, power traffic suffers from exponential power attenuation according to its propagation distance, resulting in a doubly near-far problem [5–6]. Accordingly, sensor nodes located farther away from the HAP harvest less energy. As a result, the imbalance in the amount of residual energy between sensor nodes and the inefficient use of channel resources due to the doubly near-far problem in WPSN are considered challenging issues.

To address the doubly near-far problem in WPSN, various techniques have been proposed. In [7–11], resource allocation schemes were proposed to determine the WPT and data transmission period based on the distance between the sensor node and the HAP. These schemes aim to provide fairness between sensor nodes in terms of throughput and residual energy. However, they do not consider changes in the channel state between the sensor node and the HAP, which significantly impacts the amount of energy harvested by the sensor node. Additionally, sensor nodes harvest energy and transmit data packets through a single HAP without considering the different characteristics of power and data traffic. Accordingly, some sensor nodes request power from a more distant HAP rather than the nearest one, leading to unnecessary long energy harvesting. This causes transmission delays of data packets and bottlenecks, ultimately resulting in the degradation of WPSN performance.

In this paper, we propose a Q-learning-based traffic-aware parent selection for WPSN (QTaPS), which aims to mitigate the inefficient use of channel resources and reduce the transmission delay of data traffic. QTaPS enables the sensor node to adaptively select its data parent and power parent for energy harvesting and data transmission based on their respective traffic characteristics. To this end, we define the states and actions of the two Q-learning agents for the parent selection. In addition, the rewards for power and data parent selection are defined using multiple parent selection metrics, such as hop count, distance, link quality, and traffic load, which consider the characteristics of power and data traffic. As a result, the sensor node selects the power parent and data parent based on the updated Q-values. We conducted an experimental simulation to verify the superiority of QTaPS. As a result, QTaPS achieved 2.85% and 20.94% higher aggregate throughput and 10.43% and 39.30% lower average end-to-end delay compared to the existing parent selection schemes, respectively.

The rest of this paper is organized as follows. Section II presents a system model for multi-hop WPSN. Section III describes the design of QTaPS. We provide the simulation configuration and results in Section IV. Finally, Section V concludes this paper.

II. System Model

A. System Architecture

Figure 1 depicts the system architecture of a multi-hop WPSN consisting of multiple HAPs and sensor nodes. We assume that HAPs and sensor nodes can only select a HAP that supports both WPT and data communication as their parent (i.e., next hop). Additionally, we assume that the sensor node selects two independent parents (i.e., power parent and data parent) based on power traffic and data traffic, each with different characteristics. On the other hand, HAP does not require energy harvesting, so it only selects the data parent as its parent. HAP supplies power to sensor nodes and transmits data traffic received from the sensor nodes to its data parent. It is equipped with a directional antenna for WPT and an omnidirectional antenna for transmitting data traffic. Also, one HAP can serve as both the power parent and data parent of the sensor node. In other words, the power parent and data parent of the sensor node are the same HAP. The sensor node harvests energy from its power parent and then transmits its data packet to its data parent. It operates with a limited battery capacity and is equipped with an omnidirectional antenna. Additionally, it is assumed that the WPSN adopts a carrier

Corresponding Author: ejkim32@hallym.ac.kr

sense multiple access with collision avoidance (CSMA/CA) channel access mechanism.

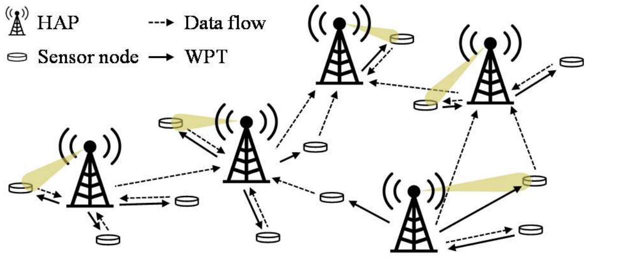

Fig. 1. System architecture of multi-hop WPSN.

B. Energy Model

In a multi-hop WPSN, the amount of energy harvested by the sensor node depends on the transmission power of the HAP, the wavelength of the RF signals, and the distance between the HAP and the sensor node. The received power of the sensor node from the HAP (i.e., power parent) (P_r) in free space can be calculated using the Friis equation as follows:

$$P_r = P_t G_t G_r \left(\frac{\lambda}{4\pi d} \right)^2 \tag{1}$$

where P_t is the transmission power of the power parent. G_t and G_r are the antenna gains of the power parent and the sensor node, respectively. λ is the wavelength of the RF signal. d is the distance between the sensor node and its power parent. Thus, the amount of energy harvested by the sensor node (P_{Rx}) for the duration of energy harvesting (T_{EH}) can be calculated as follows:

$$P_{Rx} = \eta P_r T_{EH} \tag{2}$$

where η is the energy harvesting efficiency.

The sensor node changes its state to one of four states: transmission, reception, idle, and sleep. The amount of energy consumed by the sensor node can be calculated based on its operation as follows:

$$E_{Tx} = E_{idle} T_{backoff} + E_{rx} T_{cca} + E_{tx} T_{data} + E_{rx} T_{ack} \tag{3}$$

$$E_{Rx} = E_{rx} T_{data} + E_{rx} T_{ack} \tag{4}$$

$$E_{Idle} = E_{idle} T_{idle} \tag{5}$$

$$E_{Sleep} = E_{sleep} T_{sleep} \tag{6}$$

where E_{Tx} , E_{Rx} , E_{Idle} , and E_{Sleep} are the amount of energy consumed by the sensor node during successful transmission, successful reception, idle state, and sleep state, respectively. E_{tx} , E_{rx} , E_{idle} , and E_{sleep} are the amount of energy consumed per second by the sensor node, respectively. $T_{backoff}$, T_{cca} , T_{data} , and T_{ack} refer to the backoff period, clear channel assessment (CCA) period, length of the data packet, and length of the Ack packet, respectively. T_{idle} and T_{sleep} represent the periods during which the sensor node remains in the sleep state and idle state, respectively.

The sensor node calculates its residual energy by considering the amount of energy harvested and consumed. Additionally, the sensor node requests a power supply from its power parent to harvest the energy, which is necessary for transmitting the data packet through a contention-based channel access.

III. Design of QTaPS

QTaPS is designed to enable the sensor node to adaptively select its power parent and data parent by considering parent selection metrics suitable for power and data traffic. The multi-hop WPSN can be modeled as an environment that includes HAPs, sensor nodes, and power and data traffic transmitted between sensor nodes and HAPs. The sensor node selects its power parent and data parent separately, utilizing each Q-learning agent for the power parent and data parent. All HAPs within WPSN are defined as a set of states and actions, which are expressed by $\mathbf{S} = \{s_1, s_2, \cdots, s_m\}$ and $\mathbf{A} = \{a_1, a_2, \cdots, a_m\}$ where m is the number of HAPs in the WPSN. The sets of states ($\mathbf{S}_{PP}$ and $\mathbf{S}_{DP}$) and actions ($\mathbf{A}_{PP}$ and $\mathbf{A}_{DP}$) of each agent for the power parent and data parent are defined as $\mathbf{S}_{PP} = \{s_1, s_2, \cdots, s_k\}$, $\mathbf{S}_{DP} = \{s_1, s_2, \cdots, s_k\}$, $\mathbf{A}_{PP} = \{a_1, a_2, \cdots, a_k\}$, and $\mathbf{A}_{DP} = \{a_1, a_2, \cdots, a_k\}$ where k is the number of neighbor HAPs of the sensor node. Each agent transitions from one state to another by performing one action from its set of actions. Each agent interacts with the environment by performing actions (a_t^{PP} and a_t^{DP}) in the current state at t (s_t^{PP} and s_t^{DP}), and then obtains the corresponding reward (r_t^{PP} and r_t^{DP}). t refers to the number of executions of QTaPS for each agent. The reward functions of agents for power and data parent consider the parent selection metrics (i.e., hop count, distance, link quality, and traffic load), which reflect the characteristics of power and data traffic, and are defined as follows:

$$r_t^{PP}\left(s_t^{PP}, a_t^{PP}\right) = hopCnt / \left(d \times tl\right) \tag{7}$$

$$r_t^{DP}\left(s_t^{DP}, a_t^{DP}\right) = etx / \left(hopCnt \times tl\right) \tag{8}$$

where $hopCnt$, tl , and etx are the hop count, traffic load, and link quality of the HAP selected by the action of sensor node, respectively. The hop count refers to the number of intermediate devices in the routing path from the selected HAP to the root. The traffic load refers to the total amount of data and power traffic of the selected HAP. The link quality refers to the expected number of transmissions required by the sensor node to successfully transmit a data packet to the selected HAP. Each agent maintains a Q-table to select the best parent, in which the action value function $Q(s_t, a_t)$ returns the expected sum of immediate and future rewards when an action a_t is selected in state s_t at t . The action value functions of two agents are defined as follows:

$$\begin{aligned} &Q\left(s_t^{PP}, a_t^{PP}\right) \\ &= (1-\alpha) Q\left(s_t^{PP}, a_t^{PP}\right) + \alpha \left\{ r_t^{PP} + \gamma \max Q\left(s_{t+1}^{PP}, a_t^{PP}\right) \right\} \end{aligned} \tag{9}$$

$$Q\left(s_t^{DP}, a_t^{DP}\right) = (1-\alpha) Q\left(s_t^{DP}, a_t^{DP}\right) + \alpha\left\{r_t^{DP} + \gamma \max Q\left(s_{t+1}^{DP}, a_t^{DP}\right)\right\} \quad (10)$$

where α is the learning rate and γ is the discount factor for future rewards. Each agent selects the action with the highest Q-value in the Q-table for each state and receives the reward. Then, it updates the Q-values for each state and transitions to a new state. Therefore, the sensor node independently selects the power and data parent based on Q-tables updated by the two agents, taking into account the different characteristics of power and data traffic in the dynamic network environment.

IV. Performance Evaluation

A. Simulation Configuration

We conducted an experimental simulation to verify the performance of QTaPS under the slotted CSMA/CA in IEEE 802.15.4 using the MATLAB simulator. The simulation results of QTaPS were compared with those of the object function zero (OF0) [12] and the minimum rank with hysteresis objective function (MRHOF) [13] in terms of aggregate throughput and average end-to-end delay. OF0 and MRHOF allow the sensor node to select a single parent based on the hop count and link quality, respectively. In the simulation, a multi-hop WPSN comprised of 25 HAPs and sensor nodes is considered. The number of sensor nodes ranges from 2 to 20. HAPs are arranged in a 5×5 grid structure. The sensor nodes are randomly deployed within a 120×120 m^2 and periodically generate a data packet every 0.2 seconds. The detailed simulation parameters are listed in Table I.

TABLE I. Simulation Parameters

Parameter	Value	Parameter	Value
PHY	IEEE 802.15.4	d	0–2 m
MAC	Slotted CSMA/CA	P_t	100 mW
Number of HAPs	25	E_{tx}	20.98 mA
Number of sensor nodes	2–20	E_{rx}	17.96 mA
Data packet size	127 bytes	E_{idle}	0.001 mA
Ack packet size	5 bytes	E_{sleep}	0.001 mA
Superframe order (SO)	6	η	0.65
Beacon order (BO)	6	λ	0.125
SIFS	192 μs	α	0.8
LIFS	640 μs	γ	0.9

B. Simulation Results

Fig. 2 illustrates the aggregate throughput for varying number of sensor nodes. Overall, the aggregate throughput increases until a certain number of sensor nodes is reached and then tends to decrease as the number of sensor nodes increases. This is because an increase in the number of sensor nodes leads to higher power and data traffic within the network and intensifies competition between sensor nodes trying to access the channel. It also results in collisions and transmission delays of data packets. QTaPS exhibits a higher aggregate throughput compared with both OF0 and MRHOF. QTaPS enables the sensor node to select the HAP that is relatively close to the root and has a low traffic load among neighbor HAPs as its data parent. On the other hand, OF0 and MRHOF use a single parent for both energy harvesting and data transmission. Additionally, as OF0 performs parent selection based on the hop count, it creates a shorter path from the sensor node to the root HAP compared to MRHOF, thereby achieving higher aggregate throughput. Quantitatively, the aggregate throughput of QTaPS is 2.85% and 20.94% higher than that of OF0 and MRHOF, respectively.

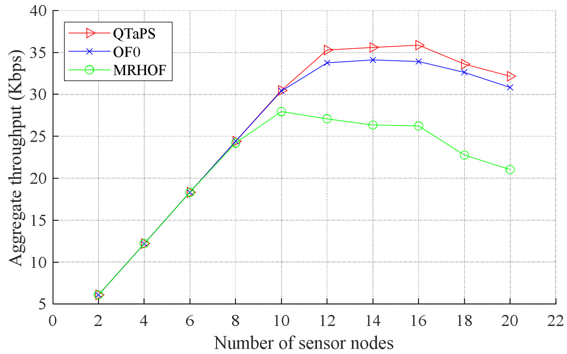

Fig. 2. Aggregate throughput.

Fig. 3 illustrates the average end-to-end delay for varying number of sensor nodes. Overall, the average end-to-end delay increases rapidly after the number of sensor nodes reaches a certain number. This is because as the number of nodes increases, both power and data traffic also increase, resulting in more collisions and transmission delays of data packets within the network. Additionally, energy harvesting and backoff delay for channel access between sensor nodes have a significant effect on end-to-end delay in WPSN. QTaPS achieves a lower average end-to-end delay than OF0 and MRHOF. This is because the sensor node using QTaPS utilizes two independent parents to harvest energy from a nearby HAP and transmit data packets to the HAP closer to the root. The closer the distance between the sensor node and its power parent, the less time the sensor node spends harvesting the required energy. In OF0 and MRHOF, sensor nodes only use a single parent regardless of the traffic type. As a result, HAPs more frequently suffer from bottlenecks due to the concentration of power and data traffic. Quantitatively, the average end-to-end delay of QTaPS is 10.43% and 39.30% lower than that of OF0 and MRHOF, respectively.

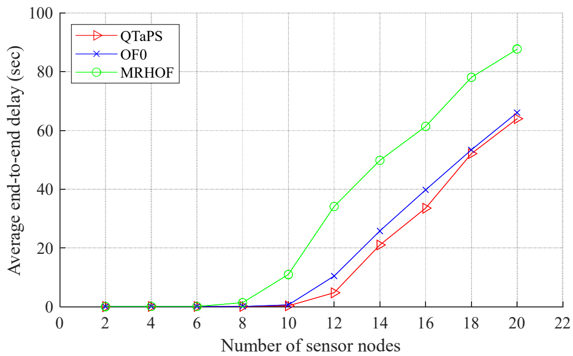

Fig. 3. Average end-to-end delay.

V. CONCLUSION

This paper presents the QTaPS for WPSNs, which aims to reduce the transmission delay of data packets and efficiently utilize channel resources by considering the characteristics of power and data traffic. QTaPS enables the sensor node to select two separate parents for data and power traffic by employing two Q-learning agents. To this end, we define the states, actions, and reward functions of two agents for the parent selection. We also consider the parent selection metrics (i.e., hop count, distance, link quality, and traffic load). We conducted an experimental simulation to evaluate the performance of the QTaPS. The simulation results demonstrated that QTaPS selects a more suitable parent for energy harvesting and data transmission compared with the existing parent selection schemes. Quantitatively, QTaPS achieved 2.85% and 20.94% higher aggregate throughput and 10.43% lower average end-to-end delay compared with OF0 and MRHOF, respectively.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2020R1I1A3052733). This work was also supported by the NRF grant funded by the Korean government (MSIT) (NRF-2021R1C1C2095696).

REFERENCES

[1] S. M. A. Huda, M. Y. Arafat, and S. Moh, "Wireless power transfer in wirelessly powered sensor networks: A review of recent progress", Sensors, vol. 22, no. 8, pp. 2952:1–2952:34, Apr. 2022.

[2] D. Niyato, D. I. Kim, M. Maso, and Z. Han, "Wireless powered communication networks: Research directions and technological approaches," IEEE Wireless Commun., vol. 24, no. 6, pp. 88–97, Dec. 2017.

[3] C. Shao, H. Roh, and W. Lee, "Next-generation RF-powered networks for Internet of Things: Architecture and research perspectives," J. Netw. Comput. Appl., vol. 123, pp. 23–31, Dec. 2018.

[4] B. Clerckx, R. Zhang, R. Schober, D. W. K. Ng, D. I. Kim, and H. V. Poor, "Fundamentals of wireless information and power transfer: From RF energy harvester models to signal and system designs," IEEE J. Sel. Areas Commun., vol. 37, no. 1, pp. 4–33, Jan. 2018.

[5] S. Bi, C. K. Ho, and R. Zhang, "Wireless powered communication: Opportunities and challenges," IEEE Commun. Mag., vol. 53, no. 4, pp. 117–125, Apr. 2015.

[6] C. He, J. Liang, G. Qian, C. Guo, and D. Feng, "Optimal time allocation in multi-cell wireless powered communication networks," IEEE Access, vol. 7, pp. 26519–26526, Nov. 2019.

[7] M. Lei, X. Zhang, H. Ding, and B. Yu, "Fairness-aware resource allocation in multi-hop wireless powered communication networks with user cooperation," Sensors, vol. 18, no. 6, pp. 1890:1–1890:19, Jun. 2018.

[8] C. Zhang, P. Zhang, and W. Zhang, "Cluster cooperation in wireless-powered sensor networks: Modeling and performance analysis," Sensors, vol. 17, no. 10, pp. 2215:1–2215:19, Sep. 2017.

[9] L. Ji and S. Guo, "Energy-efficient cooperative resource allocation in wireless powered mobile edge computing," IEEE Internet Things J., vol. 6, no. 3, pp. 4744–4754, Jun. 2019.

[10] M. Li, C.-C. Fang, and H.-W. Ferng, "On-demand energy transfer and energy-aware polling-based MAC for wireless powered sensor networks," Sensors, vol. 22, no. 7, pp. 2476:1–2476:15, Mar. 2022.

[11] H. Ju and R. Zhang, "Throughput maximization in wireless powered communication networks," IEEE Trans. Wireless Commun., vol. 13, no. 1, pp. 418–428, Jan. 2014.

[12] P. Thubert, Objective Function Zero for the Routing Protocol for LowPower and Lossy Networks (RPL), document RFC 6552, IETF, Fremont, CA, USA, 2012.

[13] O. Gnawali and P. Levis, The Minimum Rank with Hysteresis Objective Function, document RFC 6719, IETF, Fremont, CA, USA, 2012.