Journal of information and communication convergence engineering
http://jicce.org

J. Inf. Commun. Converg. Eng. 19(1): 1-7, Mar. 2021 Regular paper

Performance Enhancement of CSMA/CA MAC Protocol Based on Reinforcement Learning

Tae-Wook Kim and Gyung-Ho Hwang*, *Member*, *KIICE*

Department of Computer Engineering, Hanbat National University, Daejeon, 34158 Republic of Korea

Abstract

Reinforcement learning is an area of machine learning that studies how an intelligent agent takes actions in a given environment to maximize the cumulative reward. In this paper, we propose a new MAC protocol based on the Q-learning technique of reinforcement learning to improve the performance of the IEEE 802.11 wireless LAN CSMA/CA MAC protocol. Furthermore, the operation of each access point (AP) and station is proposed. The AP adjusts the value of the contention window (CW), which is the range for determining the backoff number of the station, according to the wireless traffic load. The station improves the performance by selecting an optimal backoff number with the lowest packet collision rate and the highest transmission success rate through Q-learning within the CW value transmitted from the AP. The result of the performance evaluation through computer simulations showed that the proposed scheme has a higher throughput than that of the existing CSMA/CA scheme.

Index Terms: CSMA/CA, IEEE 802.11, Q-Learning, Reinforcement Learning, Wireless LANs

I. INTRODUCTION

Carrier sense multiple access with collision avoidance (CSMA/CA) is a distributed medium access control protocol that is widely used in the IEEE wireless local area network (LAN) standard [1]. Several studies have been conducted to improve the performance of this protocol. A recent one was based on an optimal contention window (CW) value according to the traffic conditions of the station.

Reinforcement learning is an area of machine learning technology that studies how an agent recognizes the current state, takes actions, and receives rewards accordingly. In the reinforcement learning algorithm, the agent finds a policy that selects an action to maximize the rewards that it will receive in the future.

In this paper, we propose an algorithm that uses the Q-learning technique of reinforcement learning to select a backoff number within the CW provided by the access point (AP). Through reinforcement learning, the stations can continue to select a backoff number with a high transmission success rate. This ensures that the stations select different backoff numbers, thus resulting in a situation where no colliding packets are transmitted by the stations. The station selects a backoff number with the maximum Q-value and constantly updates the Q-value of each backoff number by applying a reward according to the packet transmission result of the selected backoff number.

Unlike in the existing CSMA/CA method, the CW value does not increase when the transmitted packets collide and the AP broadcasts the CW value that is commonly used by the stations. The station selects a backoff number to transmit a packet through reinforcement learning, and the AP selects an appropriate CW value according to the current traffic situation.

The rest of this paper is structured as follows. In Section II, previous studies and related works are presented. The

Received 04 January 2021, **Revised** 13 February 2021, **Accepted** 17 February 2021
***Corresponding Author** Gyung-Ho Hwang (E-mail: gabriel@hanbat.ac.kr, Tel: +82-42-821-1751)
Department of Computer Engineering, Hanbat National University, Daejeon 34158, Republic of Korea.

Open Access https://doi.org/10.6109/jicce.2021.19.1.1 print ISSN: 2234-8255 online ISSN: 2234-8883

IEEE 802.11 wireless LAN CSMA/CA MAC protocol and the reinforcement learning technique are also described in this section, along with the previous schemes for improving the performance of the MAC protocol using reinforcement learning. In Section III, we propose a method for improving the performance of the CSMA/CA protocol using reinforcement learning. In Section IV, we compare and analyze the performances of the existing CSMA/CA protocol and the proposed method according to the simulation results. Finally, we conclude the paper in Section V.

II. RELATED WORKS

A. IEEE 802.11 Wireless LAN CSMA/CA MAC Protocol

As shown in Fig. 1, the CSMA/CA method detects the channel state and checks for the presence of a data packet to be transmitted by a station. If the channel is idle for a distributed coordination function interframe space (DIFS) time, a data packet is immediately transmitted. If the channel is busy, the station waits until the channel becomes idle. When the station detects an idle channel for a DIFS time, the station performs the backoff procedure. The backoff number is randomly selected between 0 and the CW value. If the channel is idle during a slot time, the backoff counter is decreased by 1. The station sends a packet when the backoff counter reaches 0. If another station transmits a packet during the backoff procedure, the station stops the backoff procedure and waits until the channel is idle. The station that receives the packet transmits an acknowledgment (ACK) packet after a short interframe space (SIFS) time, and the station receiving the ACK packet assumes that the sent packet has been successfully transmitted. If the ACK packet is not received within a certain time, the station that transmitted the packet regards it as a transmission failure, reconfigures the backoff number, and retransmits the packet. When retransmitting, the station doubles the CW value to reduce the probability of selecting the same backoff number among stations to prevent collisions. The maximum value of the CW is predefined, and if packet transmission is successful, the CW value is set back to the minimum value.

In the CSMA/CA protocol, if the number of stations and transmitted packets increases, the probability of collisions of packets sent by the stations increases, thus resulting in a decrease in throughput. To improve the performance, we need to reduce the probability of collision. As the CW value increases, the probability that the stations will select the same backoff number decreases. Meanwhile, the number of idle slots increases and, thus, the throughput may decrease. In the existing CSMA/CA method, the CW value is doubled each time there is a collision to determine the optimal CW value. In this study, the AP broadcasts the CW value to the stations through a broadcasting beacon frame and each station chooses a backoff number to reduce collisions.

B. Reinforcement Learning

An overview of reinforcement learning is shown in Fig. 2. The agent determines the task that needs to be performed according to the surrounding conditions. Furthermore, the agent defines the states describing the environment and obtains a reward from the environment according to its action. The goal of reinforcement learning is to train the agent to obtain as many rewards as possible [2].

The “state” is a set of values representing the current situation, and the set of all states is called the state space. The state at a specific time is expressed as S_t.

The “action” refers to the options an agent can take. The set of actions that can be taken is called the action space, and actions in a specific state are denoted as A_t.

The “reward” is the resultant value of an action taken in a certain state. The larger the reward is, the better it is, and the agent constantly aims to maximize the cumulative rewards. The reward is the most important factor in reinforcement learning, and even in the same algorithm, the way the reward function is defined has a significant impact on the performance.

Reinforcement learning is largely classified into model-based algorithm (MBA) and model-free algorithm (MFA). The MBA is an algorithmic model that conducts learning under the assumption that all explanations of the environ-

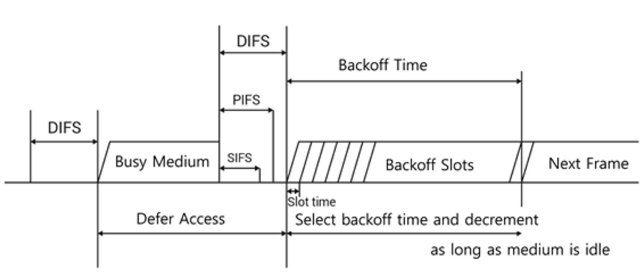

Fig. 1. Illustration of the IEEE 802.11 wireless LAN CSMA/CA MAC protocol.

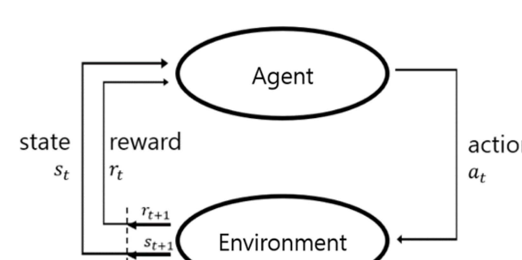

Fig. 2. Overview of reinforcement learning.

ment are known. Therefore, the state and action required for an optimal result can be determined.

Unlike the MBA, the MFA initiates learning without any information about the environment. Because the agent has no information about the environment, it manually sets the next state or reward. The agent learns through a trial-and-error method during exploration. Through this process, the future reward is maximized. An example of an MFA is Q-learning [3-4]. Q-learning is a method that uses a Markov decision process. Like in general reinforcement learning, Q-learning derives the optimal result through learning; however, it uses a Q-value for each (state and action) pair. The highest Q-value among the actions is chosen for the state. The Q-value is updated as shown in (1).

$$Q(S_t, A_t) \leftarrow (1-\alpha) \cdot Q(S_t, A_t) + \alpha \cdot \{r_t + \gamma \cdot \max_a(S_{t+1}, A)\} \tag{1}$$

where α represents the learning rate, r_t represents the reward, and γ represents the discount rate. Here, the learning rate refers to the rate at which the newly learned value will be used and the discount rate refers to the importance of the current reward over the future reward.

C. Q-Learning-Based ALOHA

The slotted ALOHA MAC protocol, which uses reinforcement learning, was proposed in [5]. This protocol aims to find the slot where each node can transmit a packet without collisions. All the nodes have a Q-value for each slot. The Q-value is defined in (2).

$$Q_{t+1} = Q_t(i,k) + \alpha(r - Q_t(i,k)), \tag{2}$$

where α is the learning rate and r is the current reward. Here, i is the node number, k is the slot number, and t is the order of the frames. If the transmission succeeds, the reward will be increased by a value of 1, and if the transmission fails, a value of -1 is added to the reward. The node selects the slot with the highest Q-value.

D. Reinforcement Learning Dynamic Frame Slotted ALOHA

A radiofrequency identification (RFID) anti-collision algorithm based on reinforcement learning was proposed in [6]. The reinforcement learning dynamic frame slotted ALOHA (RL-DFSA) algorithm improves the performance of the Q-algorithm, which is itself based on the DFSA method, through Q-learning. The states are defined as the number of slots where collisions occur in one frame, and the actions are defined as 11 different frame sizes. Four different rewards are used by defining the ratio of the slots where packet collisions occur among all slots of the frame.

In RL-DFSA, the actions utilize the results of Dhakal and Shin's scheme [7], wherein the most appropriate frame size is 1.46 × the number of remaining tags. In the performance evaluation, the number of tags was set to 1000, and the collision slots were represented by 500 states. With the use of the RL-DFSA method, the transmission completion time of the RFID anti-collision algorithm was reduced compared with that of the existing RFID anti-collision MAC protocol.

III. REINFORCEMENT LEARNING-BASED CSMA/CA MAC PROTOCOL

In this paper, we propose a new wireless MAC protocol by applying reinforcement learning to the wireless LAN CSMA/CA standard. In the existing IEEE 802.11 wireless LAN CSMA/CA MAC protocol, all stations change the CW value according to the failure or success of the packet transmission. However, in the proposed method, all stations use the same CW and select one backoff number from 1 to the CW value according to the Q-learning algorithm. To determine the appropriate CW value in the given traffic conditions, the AP changes the CW value according to the number of consecutive successes or collisions of the transmitted packets. The CW value is transmitted to all stations by the AP via a beacon frame.

The proposed MAC protocol requires an operation algorithm for each AP and station. The AP adjusts the CW value according to the channel condition, and the station finds a backoff number that can maximize the transmission success rate through a Q-learning technique within the received CW value.

The AP operation is shown in Fig. 3. The AP checks the presence of packet transmission in every slot, and if the channel is idle, the AP does nothing. If a station in the slot transmits a packet, the AP receives that packet and checks whether it has been properly received. The AP continuously updates two parameter values, namely, N_S and N_F, where N_S is the number of successive successful packet receptions and N_F is the number of successive failed packet receptions due to collision. If the AP receives the packet correctly, it increases the N_S value by 1 and initializes the N_F value to 0. If the contents of the packet are broken owing to a collision, the AP increases the N_F value by 1 and sets the N_S value to 0. When the N_S and N_F values exceed the predefined thresholds, the AP selects a new CW value and broadcasts it to all the stations.

If N_S is greater than TH_S, i.e., the successive successful packet reception threshold, the new CW_{new} is reduced to a value less than that of the existing CW value. TH_S is determined by the ratio of the current CW value, as in (3). The new CW_{new} is determined by multiplying the current CW value by M_S, as in (4). M_S must be a number greater than 0

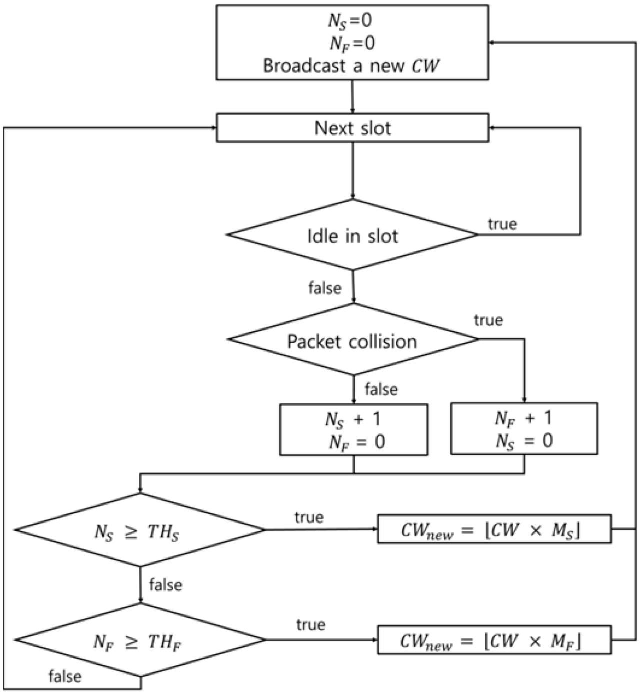

Fig. 3. Algorithm for AP to adjust CW

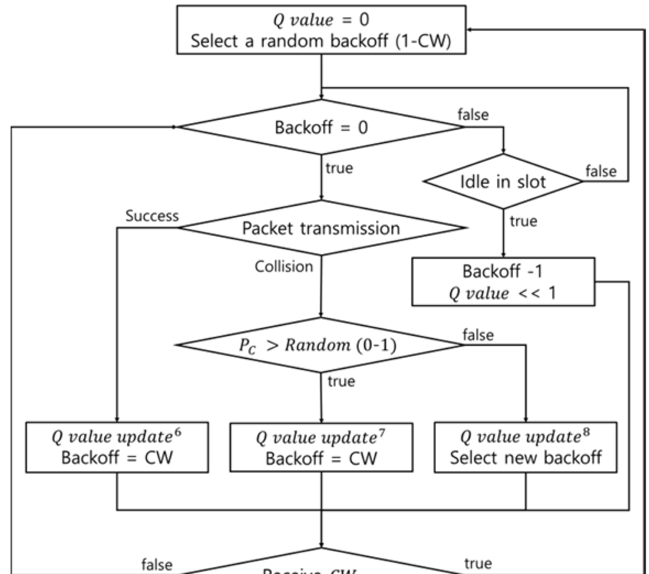

Fig. 4. Algorithm for a station to determine its own backoff number using reinforcement learning.

and less than 1. The $\lfloor v \rfloor$ denotes the maximum integer value less than v.

$$TH_S = \lfloor CW \times A \rfloor \tag{3}$$

$$CW_{new} = \lfloor CW \times M_s \rfloor \tag{4}$$

If the AP receives the corrupted packet continuously, and the N_F is greater than TH_F, which is the predefined successive packet collisions threshold, the CW_{new} value is obtained by multiplying the existing CW by M_F as (5). TH_F should use a number greater than 0, and M_F should use a number greater than 1.

$$CW_{new} = \lfloor CW \times M_F \rfloor \tag{5}$$

The station operation of the proposed CSMA/CA protocol based on reinforcement learning is shown in Fig. 4. The station randomly selects a backoff number in the range from 1 to the CW value, which is received from the AP, and generates a Q-value array of size CW. When the station receives a new CW value from the AP, all Q-values are initialized to 0. If no station in the current slot is transmitting, the backoff counter is reduced by 1, and the index and value matching the Q-value array are circularly shifted by 1. If there is only one station transmitting in the current slot and the packet transmission is successful, the corresponding station updates the Q-value, as shown in (6), and the next backoff number is determined as the CW value to maintain the selected backoff number. Q′(S) is a Q-value for the current backoff number, and R_S is the reward for packet success.

$$Q(S) = \gamma \times Q'(S) + R_S \tag{6}$$

When transmitted packets collide, the station generates a random number between 0 and 1. If this number is less than P_C, i.e., a predefined probability to maintain the backoff number, the station updates the Q-value, as shown in (7). The backoff number is set to the CW value, which means that the previously selected backoff number is used again. R_{FC} is a reward when the selected backoff number is reused, even if the packet with the backoff number collides.

$$Q(S) = \gamma \times Q'(S) + R_{FC} \tag{7}$$

If the generated random number is greater than P_C, the Q-value is updated using (8), and a new backoff number is selected according to the Q-value of each backoff number. The station chooses the backoff number with the largest Q-value, except for the previously selected backoff number. If there are multiple backoff numbers with the largest Q-value, the station randomly selects one value. R_{FN} is the reward when a transmitted packet collides and a new backoff number is selected.

$$Q(S) = \gamma \times Q'(S) + R_{FN} \tag{8}$$

IV. SIMULATION RESULTS

A. Simulation Environments

A simulation was performed using MATLAB to verify the performance of the proposed CSMA/CA protocol using reinforcement learning. It was assumed that the station always contained packets that needed to be transmitted. The number of stations was considered to be 100, which indicates a dense traffic environment. The packets transmitted from the station to the AP were used, and a packet collision was considered to have occurred when two more stations selected the same backoff number and transmitted packets simultaneously. Packet errors due to wireless channel interference were not considered.

The simulation parameters used in this study are listed in Table 1. The performance was evaluated using several values of the variables used in the proposed MAC protocol.

The performance of the existing CSMA/CA wireless LAN standard scheme was compared with that of the proposed scheme with respect to the throughput. The throughput is defined in (9). The total time was the duration when the simulation was performed, and the number of successes indicated the total number of transmitted packets that were successfully received for the simulation time. The simulation was performed for 10 s, and the results of 100 different attempts were averaged to show the performance [8].

$$Throughput = \frac{Number\ of\ Successes \times Data\ Size}{Total\ Time} \quad (9)$$

B. Performance Evaluations

The performances of the IEEE 802.11 wireless LAN CSMA/CA MAC protocol and the proposed reinforcement learning-based MAC protocol were compared and the results are shown in Fig. 5. The initial CW values for both the standard scheme and the proposed scheme were set to 15 and 31, respectively. Other parameters were set as follows: TH_S = 2, TH_F = 5, M_S = 0.6, M_F = 2.0, and P_C = 0.3. The performances varied according to the initial value in the conventional CSMA/CA method, and the throughput decreased with the increase in the number of stations, as expected. On the other hand, in the case of the proposed MAC protocol, there was no significant change in performance even if the initial CW value was set differently and the same throughput s maintained as the number of stations increases.

Table 1. Simulation parameters

Parameters	Value
Transmitted data size	8,000 bits
PHY data rate	11 Mbps
Data transmission time	727 µs
SIFS	10 µs
DIFS	50 µs
Slot time	20 µs
Discount rate, γ	0.9
R_S	3
R_{FC}	1
R_{FN}	-1

Fig. 6 shows a comparison of the throughput performance according to the threshold variable TH_F, which was used to increase the CW size for consecutive packet collisions. The performance was evaluated for TH_F values of 3, 5, 7, and 9. The values of the variables TH_S, M_S, M_F, and P_C were set to 2.0, 0.6, 2.0, and 0.3, respectively.

When TH_F was 5, the throughput value was higher than that of other values of TH_F. The value of the throughput

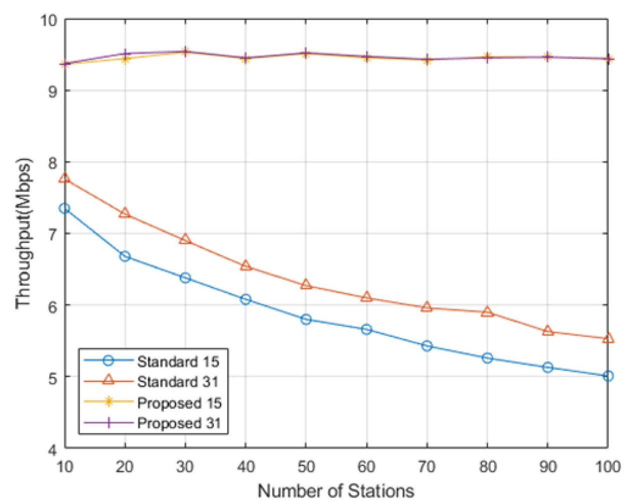

Fig. 5. Throughput comparison of proposed method and standard scheme when the initial CW is set to 15 and 31 respectively.

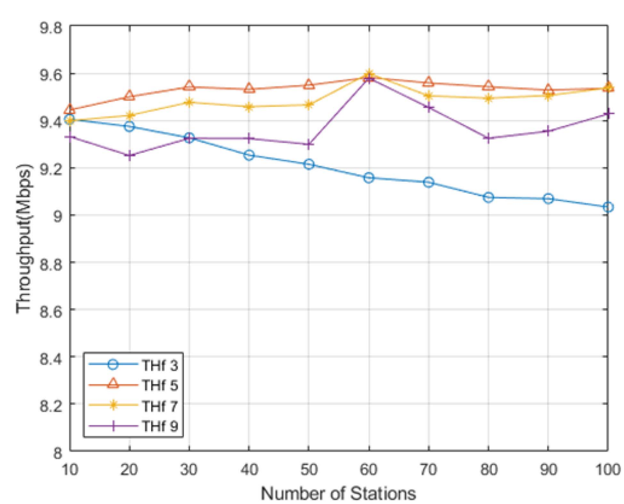

Fig. 6. Throughput performance for different values of TH_F.

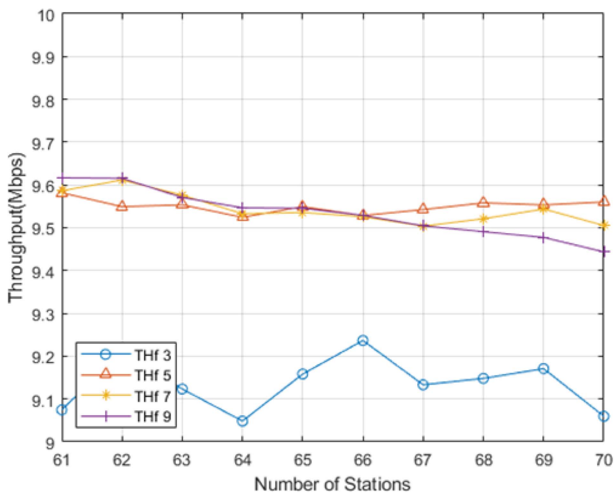

Fig. 7. Throughput performance for different values of TH_F when the number of stations was 61 to 70.

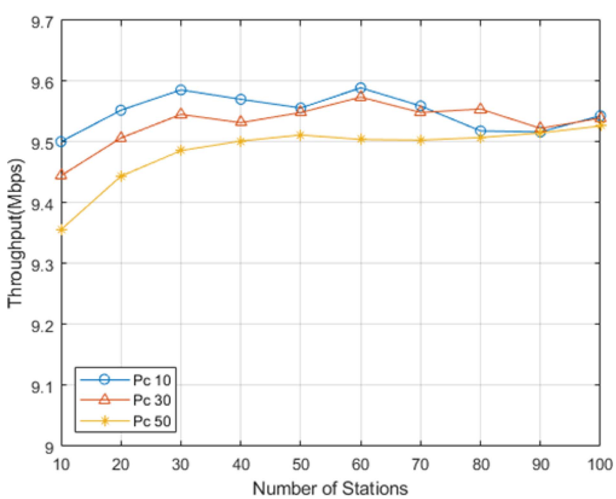

Fig. 8. Throughput performance for different values of P_C.

when TH_F was 9 was lower than that when TH_F was 5 or 7. However, when the number of stations was 60, the throughput was almost the same as that when TH_F was 5 or 7. The precision of the system can be verified in Fig. 7, which shows the throughput values when the number of stations changed from 61 to 70. From these results, it can be inferred that the parameters in the proposed algorithm are optimal for specific environments.

Fig. 8 shows the throughput performance with the change in P_C, which represents the probability that the selected backoff number is maintained when a transmitted packet collides in a station operation. The variables TH_S, TH_F, M_S, and M_F were set to 2.0, 0.5, 0.6, and 2.0, respectively. When the probability P_C was set to 0.1, which indicated that 10% of the collided stations maintained the previously selected backoff number, the performance was better than that in the case when P_C was 0.3 or 0.5. However, the best P_C value for the performance was found to vary depending on the number of stations.

V. DISCUSSION AND CONCLUSIONS

In this paper, we proposed an algorithm based on reinforcement learning to improve the performance of the IEEE 802.11 wireless LAN CSMA/CA MAC protocol. The AP continuously monitors the status of packets transmitted from the stations and informs the stations of the changes in CW according to the successive number of transmission successes or failures. The station selects a backoff number with the highest packet transmission success rate within a given CW using the Q-learning technique. From the simulation results, it can be inferred that the throughput of the proposed method is higher than that of the existing CSMA/CA method. Furthermore, the throughput of the proposed method is maintained even when the number of stations increases.

In future work, we will examine the performance variation according to the various parameters of the proposed method and study a method to select the optimal parameter value by using reinforcement learning.

ACKNOWLEDGMENTS

This research was supported by the research fund of Hanbat National University in 2017.

REFERENCES

[1] "IEEE Standard for Information technology-Telecommunications and information exchange between systems Local and metropolitan area networks-Specific requirements- Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," in IEEE Std 802.11-2016, Dec. 2016. DOI: 10.1109/IEEESTD.2016.7786995.

[2] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction,* A Bradford Book, 2018. DOI: 10.5555/3312046.

[3] H. V. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proceedings of AAAI Conference on Artificial Intelligence*, pp. 2094-2100, 2016. DOI: 10.5555/3016100.3016191.

[4] S. Galzarano, A. Liotta, and G. Fortino, "QL-MAC: A Q-learning based MAC for wireless sensor networks," in *Proceedings of International Conference on Algorithms and Architectures for Parallel Processing*, pp. 267-275, 2013. DOI: 10.1007/978-3-319-03889-6_31.

[5] Y. Chu, P. D. Mitchell, and D. Grace, "ALOHA and Q-learning based medium access control for wireless sensor networks," in *Proceedings of International Symposium on Wireless Communication Systems (ISWCS)*, pp. 511-515, 2012. DOI: 10.1109/ISWCS.2012.6328420.

[6] M. Loganathan, T. Sabapathy, and M. Elshaikh, “Reinforcement learning based anti-collision algorithm for RFID systems,” *International Journal of Computing*, pp. 155-168, 2019. DOI: 10.47839/ijc.18.2.1414.

[7] S. Dhakal and S. Shin, “Precise-optimal frame length based collision reduction schemes for frame slotted Aloha RFID systems,” *KSII Transactions on Internet and Information Systems*, vol. 8, no. 1, pp. 165-182, 2014. DOI: 10.3837/tiis.2014.01.010.

[8] T. Ho and K. Chen, “Performance analysis of IEEE 802.11 CSMA/CA medium access control protocol,” in *Proceedings of PIMRC’96 - 7th International Symposium on Personal, Indoor, and Mobile Communications*, pp. 407-411, 1996. DOI: 10.1109/PIMRC.1996.567426.

Tae-Wook Kim

received his B.S. and M.S. degrees from the Department of Computer Engineering at Hanbat National University, Daejeon, Korea, in 2019 and 2021, respectively. His research interests include artificial intelligence, reinforcement learning, and wireless sensor networks.

Gyung-Ho Hwang

received his B.S., M.S., and Ph.D. degrees from the Department of Electrical Engineering and Computer Science at KAIST, Daejeon, Korea, in 1998, 2000, and 2005, respectively. He is currently working as a professor in the Department of Computer Engineering, Hanbat National University, Korea. His research interests include wireless communication protocols based on machine learning.