Reinforcement Learning-based MAC for Reconfigurable Intelligent Surface-Assisted Wireless Sensor Networks

Faisal Ahmed*, Ethungshan Shitiri†, Ho-Shin Cho†
*†Kyungpook National University, Daegu, South Korea
{ethungshan, hscho}@ee.knu.ac.kr, *faisal-24@knu.ac.kr

***Abstract*—In this short paper, a reinforcement learning based back-off mechanism is proposed for a Reconfigurable Intelligent Surface (RIS)-assisted wireless sensor network. The proposed scheme has the capability to enable the sensors to access the RIS in an interference-free manner based on the intelligently selected back-off values. One of the main features of the proposed scheme is that sensors can avoid access interference without any need of additional signaling. Simulation results demonstrate that the proposed scheme significantly achieves higher network throughput and energy efficiency compared to benchmark Binary Exponential Back-off (BEB).**

***Index Terms*—back-off, interference, medium access control, RIS, Q-learning, wireless sensor networks.**

I. Introduction

In recent years, significant research attention has devoted to explore novel wireless communication paradigms in which the implicit randomness of the wireless propagation environment are exploited in order to improve the energy efficiency of wireless networks [1]. In this regard, Reconfigurable Intelligent Surface (RIS) emerges as a promising technology, which has the capability of reconfiguring the wireless propagation environment [2]. RIS is an inexpensive electromagnetic material comprising of a large number of nearly-passive elements that are able to perform specific task based on the application requirements, such as, reflection, refraction, absorption, beamforming, etc. [3].

The inherent nearly-passive property of the RIS elements means that they have ultra-low power requirements that makes RIS an attractive technology from an energy efficiency viewpoint. RIS has the capability of amplifying and forwarding the incoming signal by using RIS resources (i.e., RIS elements) as reflector without utilizing any power amplifier [4]. More precisely, by carefully mapping the phase shifts of each reflective passive element, it is possible to constructively fuse each reflected signal at the receiver. In this way, very low energy is consumed by a RIS compared to a regular amplifier/relay transceiver [4]. Many recent studies, i.e., [4], [5] showed that properly designed phase shifts with a sufficiently large RIS can outperform relay-assisted systems in terms of energy efficiency and data rate.

Although recent research illustrates that RIS is in a position to enhance physical layer performance, such as, achievable data rate, wireless coverage, and energy efficiency significantly, studies on multiple users accessing an RIS-enabled shared medium is still in its infancy [6]. Currently, only a couple of studies have investigated RIS-assisted medium access control (MAC) [6], [7]. In [7], the authors proposed an RIS-assisted handshaking based MAC for wireless networks that can improve the signal-to-noise (SNR) ratio, decrease the transmit power, and serve more number of users. In [6], the authors discussed about three different types of AI-assisted MAC protocols for RIS-aided wireless networks, i.e., centralized AI-assisted MAC, distributed AI-assisted MAC, and hybrid AI-assisted MAC; and analyzed the performance comparison among these three types in terms of throughput and energy efficiency.

From a MAC design viewpoint, energy efficiency can be further enhanced by avoiding access interference though it is challenging to coordinate the channel access of large number of RIS-assisted sensors in terms of interference-free DATA transmission. Therefore, in this preliminary study, a RIS-assisted sensor network is considered, in addition to that, a distributed reinforcement learning (RL)-based MAC protocol is proposed to access the RIS elements in an interference-free way to increase the throughput and energy efficiency of the network.

II. System Model

A multi-sensor uplink wireless communication system is considered where an RIS that consists of N number of passive elements is equipped with a RIS controller to aid communications of I number of sensors. The RIS controller is directly connected to the sink through an independent wireless channel [7]. The sensors and sink are equipped with one antenna. It is assumed that each RIS element has discrete phase shift and constant amplitude. For simplicity, we assume that the whole RIS serve one sensor at a time and the optimal phase shift of the RIS elements are known to sensors.

Time is slotted where the slot length is the combination of DATA duration and guard time. Additionally, transmission of DATA occurs at the beginning of a slot. For simplicity, the sensors are assumed to be positioned very close to each other so that signal-to-noise-interference-ratio (SINR) between the sink and the sensors are the same [7]. According to [4], [7], the SINR from sensor i (i.e., $i \in [1, I]$) to the sink can be calculated as

$$\rho_i = \frac{P_{\text{tx},i}|H_i\phi_iG_i + h_i^*|^2}{\sigma^2 + \sum_{j=1,j\neq i}^{I} P_{\text{tx},j}|H_j\phi_jG_j + h_j^*|^2} \tag{1}$$

where $P_{\text{tx},i}$, H_i, ϕ_i, G_i, h_i^*, and σ^2 stand for the transmit power of sensor i, channel gain from sensor i to the sink via the RIS, phase shift matrix of the RIS for sensor i, channel gain from sensor i to the RIS, channel gain from sensor i to the sink, and Gaussian noise variance, respectively.

III. Proposed Scheme

The proposed scheme operation is as follows. After DATA reception, sink sends back ACK or NACK to the desired sensor using the direct sensor-sink link. In short, If sink receives more than one RIS reflected DATA at the same time slot, then, the received SINR would unable to meet the minimum threshold requirement. This is because due to the access interference at the RIS, the received SINR at the sink is much lower. In such case, sink responds back with a NACK. Otherwise, sink responds back with an ACK. Based on the received ACK or NACK, corresponding sensor decides either to perform next RIS-assisted DATA transmission or retransmission of the failed DATA.

To alleviate interference at the RIS, the sensors perform back-off using reinforcement learning (i.e., Q-learning). The advantage of Q-learning is that no additional signaling is required; thus, network overhead can effectively be minimized.

A sensor is defined as an agent of Q-learning and maintains two Q tables. One table is for state-action values and the other one is for the back-off values. The initial Q-values are set to zero and are updated according to the reward obtained after an action is performed in a state using following equation

$$Q(s_t, a_t) = (1-\alpha)Q(s_t, a_t)+\alpha[r_t+\gamma \max_{a} Q(s_{t+1}, a)], \quad (2)$$

where a_t, s_t, r_t, α, γ represent the action, state, reward, learning rate, and discount factor, respectively [8].

In this preliminary work, an agent, that is, a sensor predicts the environmental state at a time slot on the basis of received ACK or NACK at its previous time slot. Therefore, the state space can be defined as

$$S = \{\psi, \zeta\}, \quad (3)$$

where ψ and ζ represent successful and unsuccessful RIS-assisted DATA transmission of a sensor, respectively. Suppose, at time slot $t-1$, if RIS-assisted DATA transmission was successful for sensor i, then, at time slot t, the state would be, $s_{i,t} = \psi_{t-1}$, Otherwise, $s_{i,t} = \zeta_{t-1}$. Here, $s_{i,t} \in S$.

In addition to that, action space can be defined as

$$A = \{\chi, \beta\}, \quad (4)$$

where χ represents RIS-assisted DATA transmission at the current time slot (back-off value is 0) and RIS-assisted DATA transmission at a later time slot is defined by β. Suppose, at time slot t, sensor i selects the action χ_t, therefore, $a_{i,t} = \chi_t$ and $a_{i,t} \in A$.

Consequently, if the selected action for sensor i at time slot t is β_t, then, back-off value would be chosen from a separate table. The back-off values are discrete for β_t and $\beta_{\text{t}} \in \{\beta_0, \beta_1, \ldots, \beta_m\}$ where β_0 and β_m represent minimum and maximum discrete back-off value, respectively. It is worth mentioning that the range of discrete minimum and maximum back-off value is different for the above mentioned two states and for ζ state, the Q value of χ sets to -100.

Once an action is performed, by using the total reward r_t, a sensor rates its action quality. The r_t is the combination of two reward factors. The first reward factor, r_1, indicates whether or not an access interference occurred at the RIS. A negative 1 is awarded if interference occurred. Conversely, a positive 1 is rewarded. The r_1 can be written as

$$r_1 = \begin{cases} +1, & \text{if ACK is received} \\ -1, & \text{if NACK is received.} \end{cases} \quad (5)$$

The second reward factor, r_2, indicates the goodness of selected back-off value, can be written as

$$r_2 = 1 - \frac{\phi_{\text{tx}} - 1}{\phi_{\max} - 1}, \quad \phi_{\max} > 1 \quad (6)$$

where ϕ_{tx} is the number of transmission attempts and $\phi_{\max}$ is the maximum allowable number of transmission attempts. The total reward for learning the back-off value to avoid access interference at the RIS is as follows

$$r_t = \sum_{i=1}^{2} r_i, \quad (7)$$

On obtaining the r_t corresponding to a selected back-off, the sensor can update both Q tables. Thus, r_t sets the goal to determine a back-off value that is able to avoid access interference at the RIS.

According to [4], the total energy consumption of the network can be calculated as

$$\mathcal{E}_{\text{total}} = P_{\text{tx}} + P_{\text{ris}} + P_{\text{rx}}. \quad (8)$$

where P_{RIS} is the static power consumption of RIS hardware and P_{rx} is the reception power consumption of the sink.

IV. Performance Evaluation

The performances are evaluated in terms of network throughput and network energy efficiency through computer simulation using MATLAB. For performance comparison, we consider two existing schemes: RIS-assisted sensor network with Binary Exponential Back-off (BEB) mechanism [9] and a relay-based sensor network with the proposed Q learning scheme. The simulation parameters are summarized in Table I.

The two performance matrices, i.e., network throughput (η) and network energy efficiency ($\mathcal{E}_{\text{eff}}$) are defined as

$$\eta = \frac{n_{\text{R}}.l_{\text{D}}}{T} \text{ [bits/sec]} \quad (9)$$

$$\mathcal{E}_{\text{eff}} = \frac{n_{\text{R}}.l_{\text{D}}}{\mathcal{E}_{\text{total}}} \text{ [bits/J]}, \quad (10)$$

where n_{D}, l_{D}, and T represent the number of DATA successfully received at the sink, DATA packet size, and total simulation time, respectively.

Fig. 1 shows the network throughput performance varying number of sensors. It is exhibited that a higher network

TABLE I
SYSTEM PARAMETERS AND VALUES

Parameters	Values
Number of sensors	60
Average traffic load	0.5 [packets/sec]
Data packet size	1044 bits
ACK/NACK packet size	20 bits
Bit rate	250 kbps
P_{tx}	15 dBm
σ^2	-80 dBm
α	0.1
γ	0.9
Guard time	0.144 ms
Total simulation time	210 s

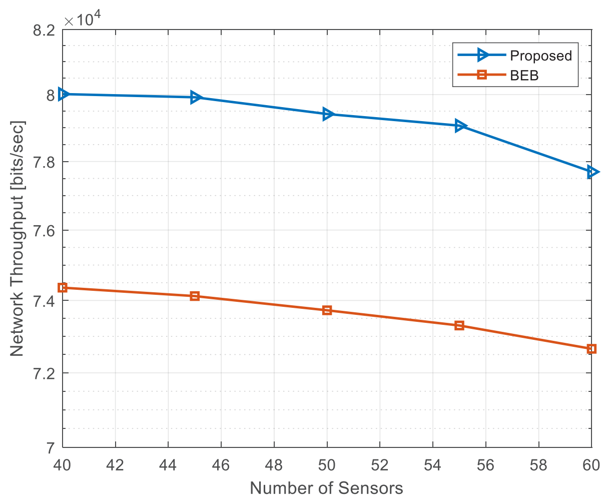

Fig. 1. Network throughput versus number of sensors.

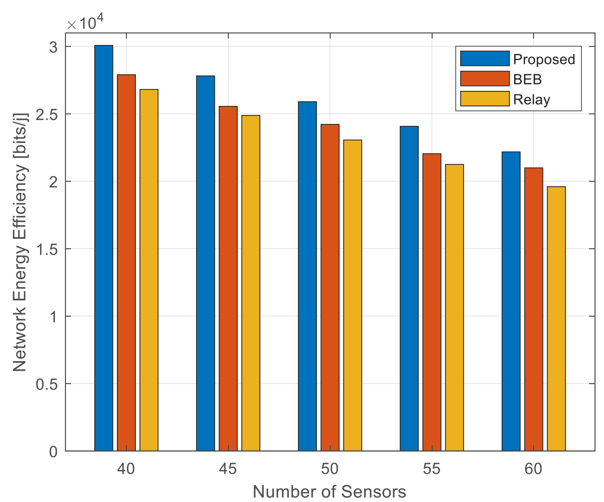

Fig. 2. Network energy efficiency versus number of sensors.

throughput is achieved that highlights the ability of the proposed initial study to adjust the back-off resulting in access interference mitigation. Thus, lower packet loses are occurred and an improved throughput performance is observed compared to BEB.

Fig. 2 shows the network energy efficiency performance varying number of sensors. It is exhibited that the proposed initial study has higher network energy efficiency compared to BEB back-off and relay. This is because the proposed initial study has effectively learned the Back-off values that can overcome access interference at the RIS and thus, reduce the total energy consumption of the network. On the other hand, the results of BEB indicate that inefficient back-off value selection results in higher packet losses, more number of retransmissions, and higher energy consumption; therefore, leading to a lower energy efficiency. As for relay, it needs reception and processing power to receive and process a DATA, respectively that results in higher energy consumption in the network degrading the network energy efficiency.

V. CONCLUSION

With the aim to mitigate RIS access interference problem, a reinforcement learning based MAC protocol is proposed, that demonstrated the capability of Q-learning to determine back-off values which actually able to improve the network's performance. Future works include developing a deep Q-learning based resource allocation algorithm which will be able to conserve more energy in RIS-assisted wireless networks.

ACKNOWLEDGMENT

This work was supported by the BK21 FOUR Project through the Ministry of Education, South Korea, under Grant 4199990113966.

REFERENCES

[1] S. Kisseleff, W. A. Martins, H. Al-Hraishawi, S. Chatzinotas, and B. Ottersten, "Reconfigurable intelligent surfaces for smart cities: Research challenges and opportunities," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1781–1797, 2020.

[2] S. Gong, X. Lu, D. T. Hoang, D. Niyato, L. Shu, D. I. Kim, and Y.-C. Liang, "Toward smart wireless communications via intelligent reflecting surfaces: A contemporary survey," *IEEE Communications Surveys Tutorials*, vol. 22, no. 4, pp. 2283–2314, 2020.

[3] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2450–2525, 2020.

[4] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 4157–4170, 2019.

[5] M. Di Renzo, K. Ntontin, J. Song, F. H. Danufane, X. Qian, F. Lazarakis, J. De Rosny, D.-T. Phan-Huy, O. Simeone, R. Zhang, M. Debbah, G. Lerosey, M. Fink, S. Tretyakov, and S. Shamai, "Reconfigurable intelligent surfaces vs. relaying: Differences, similarities, and performance comparison," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 798–807, 2020.

[6] X. Cao, B. Yang, C. Huang, C. Yuen, M. Di Renzo, Z. Han, D. Niyato, H. V. Poor, and L. Hanzo, "Ai-assisted mac for reconfigurable intelligent-surface-aided wireless networks: Challenges and opportunities," *IEEE Communications Magazine*, vol. 59, no. 6, pp. 21–27, 2021.

[7] X. Cao, B. Yang, H. Zhang, C. Huang, C. Yuen, and Z. Han, "Reconfigurable-intelligent-surface-assisted mac for wireless networks: Protocol design, analysis, and optimization," *IEEE Internet of Things Journal*, vol. 8, no. 18, pp. 14 171–14 186, 2021.

[8] Watkins, C. J.C.H, and P. Dayan, "Technical note: Q-learning," in *Machine Learning*, vol. 8, no. 3, 1992, pp. 279–292.

[9] B.-J. Kwak, N.-O. Song, and L. Miller, "Performance analysis of exponential backoff," *IEEE/ACM Transactions on Networking*, vol. 13, no. 2, pp. 343–355, 2005.