Proximal Policy Optimization for Energy-Efficient MEC Systems with STAR-RIS Assistance

Pyae Sone Aung[1], Sun Moo Kang[1], and Choong Seon Hong[1]

[1]Department of Computer Science and Engineering, Kyung Hee University, Yongin-si, Republic of Korea

***Abstract*—The growing popularity of Internet of Things (IoT) devices has led to an escalating demand for efficient data processing and transmission solutions. The concept of Mobile Edge Computing (MEC) has emerged as a potential solution to tackle these challenges by bringing computation closer to IoT devices. Nevertheless, the establishment of reliable communication connections between IoT devices and MEC servers continues to be a significant issue, especially in situations where achieving line-of-sight (LOS) conditions is troublesome. This paper studies simultaneously transmitting and reflecting reconfigurable intelligent surfaces" (STAR-RIS) to enhance communication links in MEC environments. STAR-RIS leverages the capabilities of conventional RIS to simultaneously transmit and reflect signals, thereby providing 360 °coverage. We formulate the energy minimization for all IoT devices in the STAR-RIS-assisted MEC system by jointly optimizing the energy-efficient offloading, amplitude, and phase shift coefficients of reflection and transmission of STAR-RIS elements and power control. Due to the non-convexity and coupling variables, the proximal policy optimization (PPO) technique has been adopted as a viable solution. The experimental findings presented in this study provide evidence of the efficacy of our suggested algorithm in comparison to the benchmark schemes.**

***Index Terms*—Reconfigurable intelligent surface (RIS), simultaneously transmitting and reflecting RIS (STAR-RIS), mobile edge computing (MEC), proximal policy optimization (PPO), deep reinforcement learning (DRL)**

I. Introduction

A. Background and Motivations

The rapid increment of data-intensive applications and the growing demand for real-time, low-latency services have led to the emergence of mobile edge computing (MEC) as a promising solution. MEC brings computing resources closer to the network edge, enabling faster data processing, reduced network congestion, and an improved user experience. These advantages make MEC essential for delivering fast, responsive, and efficient services in various industries. To further enhance the capabilities of MEC, the integration of reconfigurable intelligent surface (RIS) technology has gained significant attention. An RIS is a planar surface composed of a large number of passive reflecting elements, such as meta-material elements. These elements are capable of manipulating and controlling the propagation of wireless signals by dynamically adjusting their reflection properties, including phase, amplitude, and direction. The primary function of an RIS is to modify the wireless propagation environment to improve signal quality, coverage, and energy efficiency. By intelligently manipulating the reflected signals, an RIS can overcome signal attenuation, mitigate interference, and enhance the overall wireless communication performance. The RIS elements can be controlled and coordinated either by a centralized controller or by distributed algorithms that optimize signal propagation based on real-time channel conditions, user locations, and network requirements. However, there still exist limitations for conventional RIS. Since RIS focuses primarily on signal reflection, it does not have the capability to transmit signals when the transmitter and receiver are on opposite sides. Therefore, it restricts its potential for enhancing communication capacity and coverage. In order to address this, current research has focused on simultaneously transmitting and reflecting RIS (STAR-RIS), where each element is capable of both transmission and reflection of the signals concurrently, improving the quality of signals coming from both directions. Thereby, it offers 360°coverage in addition to all the advantages of RIS.

The use of machine learning techniques to solve challenging wireless communication issues has grown in popularity. Since these issues require making choices for long-term profit maximization with somewhat unpredictable outcomes, the bulk of the challenges in these sectors may be categorized as Markov Decision Process (MDP) issues. One of the prominent approaches to addressing these MDP issues is the deep reinforcement learning (DRL) approach. DRL may provide low-complexity suboptimal solutions for MDP issues in time-varying and stochastic contexts with minimal previous knowledge. As a result, we use the DRL approach in our STAR-RIS-assisted network to handle the problematic issues associated with a complicated wireless environment.

B. Research Contribution

In this paper, we investigate the DRL-based energy consumption minimization in STAR-RIS-assisted MEC systems. Since IoT devices are energy-constrained and computationally limited, certain tasks must be offloaded to the MEC server. On the other hand, since it is difficult to achieve line-of-

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2019-0-01287-005, Evolvable Deep Learning Model Generation Platform for Edge Computing), in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00207816), and in party by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2023-RS-2023-00258649) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation). *Dr. CS Hong is the corresponding author.

sight (LOS) communication between the IoT devices and the MEC server, STAR-RIS can be implemented to assist the communication links between them. The contributions of this paper fall into the following categories:

- First, we present the MEC system in which IoT devices offload computation tasks via STAR-RIS-assisted communication links.
- We formulate the energy consumption minimization problem of all IoT devices in our system by joint optimization of energy-efficient offloading, amplitude and phase shift coefficient of reflection and transmission of STAR-RIS elements, and power control.
- Since the formulated problem is non-convex and challenging to solve in polynomial time, we employ proximal policy optimization (PPO) to solve the problem.
- The effectiveness of our proposed system is validated through performance evaluation, and simulation results indicate that our proposed approach outperforms the benchmark schemes.

The remaining sections are categorized as follows. Section II lists the related works. We illustrate our system model and formulate the problem in III. Consequently, the proposed solution is given in Section IV presented. In Section V, performance evaluation is carried out to prove our proposed system's validation. Finally, Section VI brings our work to its conclusion.

II. Related Works

There have been several works on the MEC system. In [1], the authors consider computation offloading and resource allocation in wireless cellular networks with MEC. Nevertheless, these works do not consider the challenging issues of achieving LOS communication links. In [2], to provide better LOS links, the authors propose an unmanned aerial vehicle (UAV) aided MEC system where the UAV functions as both the aerial base station and MEC server. Additionally, the authors in [3] and [4] explore scenarios involving multiple UAVs for MEC applications, emphasizing energy-efficient resource allocation. It's worth noting that while the utilization of UAVs resolves the LOS issue, it also introduces supplementary energy consumption due to the inherently energy-intensive nature of UAV devices.

There are also various works on RIS-assisted communications in order to enhance communication links. In [5], the authors study energy-efficient networks with the aid of multiple RISs. The authors in [6] propose the energy-efficient system model where multiple aerial-RISs are employed to accomplish improved communication between the BS and users. In [7], the authors proposed the RIS-assisted MEC system, where RIS is implemented on UAV to support the communication links for the users to offload to the MEC server. These aforementioned works do not consider STAR-RIS, which provides a greater degree of freedom. To solve this, in [8], the authors consider the STAR-RIS-assisted MEC system. However, they do not consider DRL as a solution approach that offers advantages in solving complex problems, especially in dynamic and uncertain environments.

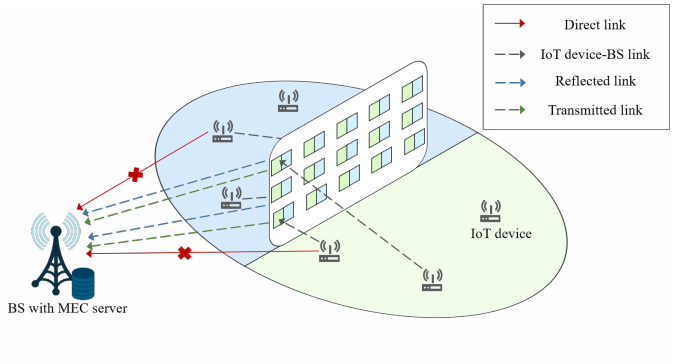

Fig. 1: System Model.

III. System Model

As illustrated in Fig. 1, we consider STAR-RIS-assisted MEC system to assist the IoT devices for the reduction of latency and energy consumption in computation tasks. In our system model, we have a BS with MEC server, a set $\mathcal{I}$ of I IoT devices, and a STAR-RIS to assist the communication due to the difficulty of achieving LOS communication links between IoT devices and BS. Given the robust computational capacity of the MEC server and the relatively modest size of the output, it is reasonable to disregard the computational time at the MEC server as well as the time required for downloading the result. The STAR-RIS consists of a set of $\mathcal{N} = \{1, 2, \ldots, N\}$ of N elements, and each element is responsible for either transmission or reflection of the incident signal towards the desired direction based on the location of the receiver. The IoT devices located in the reflection region are denoted as a set $\mathcal{R}$ of R, and the IoT devices located in the transmission region are denoted as a set $\mathcal{T}$ of T, and thereby we have total $I = R + T$ IoT devices.

A. Local Computation Model

For each IoT device, the task computed can be specified as a tuple $\{C_i, S_i, T_i^{\max}\}$, where C_i is the number of computation resources in CPU cycles required to calculate 1-bit of input data, S_i is the size of input data in bits, and $T_i^{\max}$ is the maximum tolerable delay for task completion. Since IoT devices are energy-constrained devices with limited computation capacity, it is not feasible for IoT devices to conduct the entire computation tasks locally [9]. Therefore, some of the computation tasks need to be offloaded to the MEC server. In this context, the assumption is made that the input task data bits possess bit-wise independence, enabling the breaking down into subsets of varying sizes. Consequently, these subsets may be performed in parallel by both IoT devices and the MEC server, thereby facilitating partial offloading. The variable $\alpha_i \in [0, 1]$ is used to represent the proportion of tasks to be offloaded to the MEC server, which means $(1 - \alpha_i)$ represents the proportion of tasks that are to be computed locally. The computing time required to compute the task locally can be written as

$$t_i^{\text{loc}} = \frac{(1 - \alpha_i) C_i S_i}{f_i}, \forall i \in \mathcal{I}, \tag{1}$$

where f_i is the computation frequency of IoT device i. Afterward, as in [10], the power consumption of CPU in IoT device i can be modeled as $P_i^{\text{loc}} = \kappa {f_i}^3$, where κ is the coefficient that depends on the chip design of IoT device. Therefore, the energy consumed for IoT device i to compute the task locally can be obtained as

$$E_i^{\text{loc}} = P_i^{\text{loc}} t_i^{\text{loc}} = (1-\alpha_i) C_i S_i \kappa (f_i)^2. \tag{2}$$

B. Communication Model

For the communication, we assume there is no direct link between IoT devices and BS. Therefore, the communication is assisted by the use of STAR-RIS. The indirect link consists of two parts: the IoT device to STAR-RIS link and STAR-RIS to BS link, respectively. The channel gain between IoT device r located in the reflected side and STAR-RIS can be obtained as

$$h_r = \mathbf{H}_{N,B} \mathbf{\Theta}^r \mathbf{H}_{r,N}, \forall r \in \mathcal{R}, \tag{3}$$

where $\mathbf{H}_{N,B}$ is the channel response between STAR-RIS and BS, and $\mathbf{H}_{r,N}$ is the channel response between IoT device r located in the reflected region and STAR-RIS. Similarly, the channel gain between IoT device t located in the transmitted region and STAR-RIS can be obtained as

$$h_t = \mathbf{H}_{N,B} \mathbf{\Theta}^t \mathbf{H}_{t,N}, \forall t \in \mathcal{T}, \tag{4}$$

where $\mathbf{H}_{t,N}$ is the channel response between IoT device t located in the transmitted region and STAR-RIS. The symbol $\mathbf{\Theta}^\xi, \xi \in \{r,t\}$ is the diagonal matrix of coefficients of reflection and transmission, and can be written by

$$\mathbf{\Theta}^\xi = \text{diag}(\beta_1^\xi e^{j\phi_1^\xi}, \beta_2^\xi e^{j\phi_2^\xi}, \ldots, \beta_N^\xi e^{j\phi_N^\xi}),\ \xi \in \{r,t\},\ \forall n \in \mathcal{N}, \tag{5}$$

where $(\beta_n^\xi, \xi \in \{r,t\})$ is the amplitude coefficient of STAR-RIS element n, and $(\phi_n^\xi, \xi \in \{r,t\})$ is the phase shift values of reflection and transmission. In this study, we employ the energy splitting model, as described in [11], to analyze the functioning of each STAR-RIS element. Therefore, following the principle of energy conservation, the total energy of the incident signal does not surpass the combined energies of the transmitted and reflected signals. Hence, we have

$$(\beta_n^t)^2 + (\beta_n^r)^2 = 1, \forall n \in \mathcal{N}. \tag{6}$$

We implement Rician fading model for all links: $\mathbf{H}_{N,B}$, $\mathbf{H}_{r,N}$ and $\mathbf{H}_{t,N}$, and as an example, $\mathbf{H}_{N,B}$ can be represented as

$$\mathbf{H}_{N,B} = \sqrt{\frac{K}{1+K}} \mathbf{H}_{N,B}^{\text{LOS}} + \sqrt{\frac{1}{1+K}} H_{N,B}^{\text{NLOS}}, \tag{7}$$

where K is the Rician factor, which is the ratio of the power between the LOS path and the non-LOS (NLOS) path. The received signal at the BS can be written as

$$y = \sum_{i=1}^{I} h_i x_i + \mathbf{n}_0, \forall i \in \mathcal{I}, \tag{8}$$

where $x_i = p_i s_i$ is the transmitted signal from IoT device i, which is expressed as the product of transmit power p_i and unit-power information symbol s_i. The symbol $\mathbf{n}_0$ is the additive white Gaussian noise (AWGN) with zero mean and variance σ^2. Afterward, the signal-to-noise (SNR) of IoT device r located in the reflected region can be represented as

$$\gamma_r = \frac{|h_r p_r|^2}{\sigma^2}, \forall r \in \mathcal{R}. \tag{9}$$

Likewise, the SNR of IoT device t located in the transmitted region can be represented as

$$\gamma_t = \frac{|h_t p_t|^2}{\sigma^2}, \forall t \in \mathcal{T}. \tag{10}$$

We assume that the BS's wireless frequency is divided into orthogonal sub-carriers with bandwidth W. Even if the STAR-RIS reflects/transmits all incident signals, the user may decode the received signal on the designated sub-carrier as we assume each BS-user pair is on a distinct sub-carrier. This means BS-user pairings cannot interfere [12]. Correspondingly, the uplink data rate of IoT device i can be obtained as

$$r_i = W \log_2(1+\gamma_\xi), \xi \in \{r,t\}, \forall i \in \mathcal{I}. \tag{11}$$

Given that IoT devices have limited computational capabilities and cannot execute all tasks locally, they are required to upload a certain amount of tasks to the MEC server. Therefore, the transmission delay for IoT device i to send α_i tasks for offloading may be computed as follows:

$$t_i^{\text{off}} = \frac{\alpha_i S_i}{r_i}, \forall i \in \mathcal{I}. \tag{12}$$

Then, the energy consumption of IoT device i for the task offloading can be expressed as

$$\begin{aligned} E_i^{\text{off}} = P_i^{\text{off}} t_i^{\text{off}} &= \frac{\alpha_i S_i p_i}{r_i} \\ &= \frac{\alpha_i S_i p_i}{W \log_2(1+\gamma_\xi)}, \xi \in \{r,t\}, \forall i \in \mathcal{I}. \end{aligned} \tag{13}$$

Hereby, the total energy consumption of IoT device i can be calculated as follows:

$$\begin{aligned} E_i = E_i^{\text{loc}} + E_i^{\text{off}} &= (1-\alpha_i) C_i S_i \kappa (f_i)^2 \\ &+ \frac{\alpha_i S_i p_i}{W \log_2(1+\gamma_\xi)}, \xi \in \{r,t\}, \forall i \in \mathcal{I}. \end{aligned} \tag{14}$$

C. Problem Formulation

Considering the constrained transmission time and limited data rate, our goal is to minimize the energy consumption of all IoT devices. Therefore, we formulate the joint energy-efficient offloading, amplitude, and phase shift coefficient of reflection

and transmission of STAR-RIS elements and power control problem of STAR-RIS-assisted MEC system as follows:

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{p}} \quad \sum_{i=1}^{I} E_i \tag{15a}$$

$$\text{s.t.} \quad r_i \geq r_i^{\min}, \forall i \in \mathcal{I}, \forall n \in \mathcal{N}, \tag{15b}$$

$$t_i^{\text{loc}} + t_i^{\text{off}} \leq T_i^{\max}, \forall i \in \mathcal{I}, \tag{15c}$$

$$0 \leq \phi_n^t, \phi_n^r < 2\pi, \forall n \in \mathcal{N}, \tag{15d}$$

$$(\beta_n^t)^2 + (\beta_n^r)^2 = 1, \forall n \in \mathcal{N}, \tag{15e}$$

$$0 \leq \alpha_i \leq 1, \forall i \in \mathcal{I}, , \tag{15f}$$

$$0 \leq p_i \leq p_i^{\max}, \forall i \in \mathcal{I} \tag{15g}$$

where constraint (15b) is to guarantee minimum data rate for each IoT device, constraint (15c) is the maximum tolerable latency for task completion, (15d) and (15e) are the accessible phase shift values and amplitude values for the coefficient of reflection and transmission, constraint (15f) is the offloading portion variable ranging between 0 to 1, and finally, constraint (15g) is the power budget constraint. The formulated problem is non-convex due to the coupling variables in both objective function and constraints. As a result, it is quite challenging to solve in polynomial time. Therefore, we propose one of the deep reinforcement learning approaches called proximal policy optimization (PPO) to solve our optimization problem. PPO is favored over other DRL methods for non-convex problems due to its stability, sample efficiency, and versatility. Its clipped surrogate objective ensures stable learning even in complex, high-dimensional environments while efficiently utilizing collected data. PPO's versatility and proven performance across diverse tasks make it a robust choice for tackling non-convex optimization challenges in reinforcement learning.

IV. Solution Approach

Firstly, it is essential to establish the MDP since it serves as a comprehensive framework for describing almost all DRL problems. The components included in this framework consist of the state space, action space, state transition function, reward function, and discount factor. The time horizon is divided into discrete steps as $\{1, 2, \ldots, \bar{T}\}$.

1) Agent and environment: Within our system paradigm, the BS functions as an agent responsible for making decisions. These decisions are influenced by the agent's current state and result in rewards that are provided by the system's environment, which encompasses elements such as the BS, MEC server, STAR-RIS, and channel models. Importantly, the environment undergoes dynamic changes in response to the agent's chosen actions.

2) State space: Each state $s_{\bar{t}}$ in the state space at time t can be defined as a tuple of $\{\mathbf{H}_{N,B}, \mathbf{H}_{r,N}, \mathbf{H}_{t,N}, \mathbf{\Theta}^{\xi}, C_i, S_i, T_i^{\max}, f_i, \xi \in \{r,t\}, \forall i \in \mathcal{I}, \forall n \in \mathcal{N}\}$ which contains the channel responses, coefficient matrix of reflection/transmission, information about the tasks to be computed.

3) Action space: Each action $a_{\bar{t}}$ in the action space at time t consists of decision variables of our optimization problem and can also be defined as a tuple of $\{\alpha_i, \beta_n^{\xi}, \phi_n^{\xi}, p_i, \xi \in \{r,t\}, \forall i \in \mathcal{I}, \forall n \in \mathcal{N}\}$ which contains the offloading portion, amplitude and phase shift coefficient of reflection/transmission, and transmit power. Additionally, as in [13], the amplitude and phase shift can be defined as the incremental values of the current ones and are expressed as follows:

$$\beta(\bar{t}+1) = \beta(\bar{t}) \odot \Delta\beta(\bar{t}), \tag{16}$$

$$\phi(\bar{t}+1) = \phi(\bar{t}) \odot \Delta\phi(\bar{t}), \tag{17}$$

where $\odot$ is the element-wise product, $\Delta\beta(\bar{t})$ and $\Delta\phi(\bar{t})$ represent the incremental amplitude and phase shift values, respectively.

4) Transition Function: The transition function, denoted as $P(s_{(\bar{t}+1)}|s_{\bar{t}}, a_{\bar{t}})$, describes how the environment changes from one state $s_{\bar{t}}$ to another $s_{(\bar{t}+1)}$ in response to the agent's action $a_{\bar{t}}$.

5) Reward Function: The immediate reward function $R(s_{\bar{t}}|a_{\bar{t}})$ for our problem can be defined as

$$\begin{aligned} R(s_{\bar{t}}|a_{\bar{t}}) = &\sum_{i=1}^{I} c_1 E_i + \sum_{i=1}^{I} c_2 J(r_i - r_i^{\min}) \\ &+ \sum_{i=1}^{I} c_3 J(T_i^{\max} - t_i^{\text{loc}} - t_i^{\text{off}}), \end{aligned} \tag{18}$$

where $c1, c2, c3$ are weight coefficients, and $J(x)$ is the piecewise function that can be defined as follows:

$$J(x) = \begin{cases} P^+ & \text{when } x \geq 0, \\ x, & \text{otherwise}, \end{cases} \tag{19}$$

where P^+ is the positive constant to indicate revenue.

6) Discount Factor: The discount factor η measures how much an agent prefers future benefits above immediate rewards. A large discount factor implies that the agent values long-term gains, prompting it to maximize cumulative earnings over time. However, a low discount factor helps the agent prioritize quick gains.

The functioning of our PPO method involves employing an agent, which comprises two components, namely an actor and a critic. This agent resides at the BS alongside the MEC server. The actor, as represented by the policy network $\pi_\theta(a_t|s_t)$, makes decisions on actions by considering the present observable states, with the objective of maximizing the expected cumulative rewards. The critic, represented by the value network, evaluates the quality of the states observed by the actor, providing feedback on the expected long-term rewards associated with those states. These components work together to enable the agent to adapt and update its policy in order to enhance decision-making over time and maximize cumulative discounted reward, which is defined as

$$Q(s_{\bar{t}}, a_{\bar{t}}) = \mathbb{E}\left[\sum_{\bar{t}=1}^{\bar{T}} \eta_t R(s_{\bar{t}}|a_{\bar{t}})\right]. \tag{20}$$

Algorithm 1 PPO-based STAR-RIS-assisted MEC system

for iteration=$1, 2, \ldots$ **do**
  for actor=$1, 2, \ldots$ **do**
    Collect states, actions, transition probabilities, immediate rewards from a collection of trajectories and execute old policy $\pi_{\theta_{\text{old}}}$ for time $\bar{T}$ in the environment
    Calculate generalized advantage estimators $A_1, \ldots, A_{\bar{T}}$
  end for
  Calculate $L^V(\varphi)$
  Train the actor and get optimal surrogate function $L(\theta)$
  Apply stochastic gradient descent to update φ
  Substitute θ_{old} with θ and achieve new policy π_θ
end for

The critic consists of the advantage function as follows:

$$A_{\bar{t}} = Q(s_{\bar{t}}, a_{\bar{t}}) - V_\varphi(s_{\bar{t}}), \tag{21}$$

which provides an estimate of the advantages or disadvantages of taking a specific action in a particular state, which is essential for guiding the agent's policy updates. Here, $V_\varphi(s_{\bar{t}})$ is the baseline estimate value function. Here, as in [14], we apply a generalized advantage estimator (GAE) to calculate the advantage function as

$$A_{\bar{t}} = \varepsilon_{\bar{t}} + (\eta\mu)\varepsilon_{\bar{t}+1} + \cdots + (\eta\mu)^{\bar{T}-\bar{t}+1}\varepsilon_{\bar{T}-1}, \tag{22}$$

where

$$\varepsilon_{\bar{t}} = R(s_{\bar{t}}|a_{\bar{t}}) + \eta V_\varphi(s_{\bar{t}+1}) - V_\varphi(s_{\bar{t}}), \tag{23}$$

and μ is the parameter for GAE. Afterward, the loss function derived from the temporal-difference error generated by the critic network can be expressed as

$$L^V(\varphi) = \mathbb{E}[|V_\varphi^{\text{target}} - V_\varphi(s_{\bar{t}})|], \tag{24}$$

where $V_\varphi^{\text{target}} = R(s_{\bar{t}+1}|a_{\bar{t}+1}) + \eta V_\varphi(s_{\bar{t}+1})$ [15]. PPO aims to learn an optimal policy that maximizes cumulative environment rewards. With clipping parameter c, and probability ratio $p_{\bar{t}} = \frac{\pi_\theta(a_{\bar{t}}|s_{\bar{t}})}{\pi_{\theta_{\text{old}}}(a_{\bar{t}}|s_{\bar{t}})}$, the surrogate objective function of PPO can be obtained as follows:

$$L(\theta) = \mathbb{E}[\min(p_{\bar{t}}(\theta)A_{\bar{t}}, \text{clip}(p_{\bar{t}}(\theta), 1-c, 1+c)A_{\bar{t}})]. \tag{25}$$

The purpose of the clipping parameter c is to prevent too large policy updates, which might cause instability or divergence in the learning process. In order to balance exploration and exploitation, the probability ratio $p_{\bar{t}}(\theta)$ is crucial. It helps control the size of policy updates by ensuring that they are neither too aggressive nor overly cautious. Algorithm 1 depicts the comprehensive algorithm describing the way PPO operates in our proposed system.

V. Performance Evaluation

To demonstrate the efficacy of our proposed algorithm for the STAR-RIS-assisted MEC system, we conduct numerical analysis. We employ 20 IoT devices, 10 in the reflection region and 10 in the transmission region, with STAR-RIS at the center. We performed the simulation utilizing the Unity ML-Agents toolbox in the Unity Engine [16]. The configurations for simulation are displayed in Table I. For the benchmark schemes, we compare our proposed algorithm with 1) Conventional RIS, where conventional RIS is implemented in the center instead of STAR-RIS, and 2) Random, where the amplitudes and phase shift coefficients of reflection and transmission of STAR-RIS elements are set to random values. Initially, Fig. 2 illustrates the convergence of the PPO algorithm we have proposed. Even though there is some instability in the early steps due to exploration, our algorithm converges at approximately 60,000-time steps.

TABLE I: Simulation parameters

Parameter	Value
σ^2	-174 dBm
W	1 MHz
K	3
c	0.2
S_i	[10, 50] Mbits
C_i	500 cycles
f_i	[0.5, 3] MHz
η	0.9
Learning rate	0.001
Mini batch size	16
No. of iterations	2,000
T	100,000

Fig. 3 demonstrates how total energy consumption increases with the increase in input data size. As depicted in the figure, the scenario with random amplitude and phase shift values grows exponentially with the increase in input data size, whereas our proposed algorithm and conventional RIS scheme grow progressively. Our proposed method demonstrates superior performance compared to other algorithms owing to its ability to adaptively adjust the values of amplitudes and phase shift coefficients of reflection and transmission for STAR-RIS elements from IoT devices in respective regions.

Next, Fig. 4 illustrates how total energy consumption decreases with the increase in the number of elements. In all instances, a reduction in the number of elements results in a corresponding drop in overall energy usage. Nevertheless, when considering a range of 20 to 25 elements, the reduction in energy consumption is not notably substantial. Hence, it is important to choose the optimal quantity of elements that are relevant to our system model. Our method demonstrates superior performance compared to other benchmark systems.

VI. Conclusion

In summary, this paper presents an investigation of the PPO-based STAR-RIS-assisted MEC system. To minimize the energy consumption of all IoT devices while jointly optimizing the energy-efficient offloading, amplitude, and phase shift coefficients of reflection and transmission of STAR-RIS elements and power control, we formulate our optimization problem. Given the inherent non-convexity and computational complexity of the given issue, we have chosen to use the PPO algorithm. This decision is motivated by the algorithm's

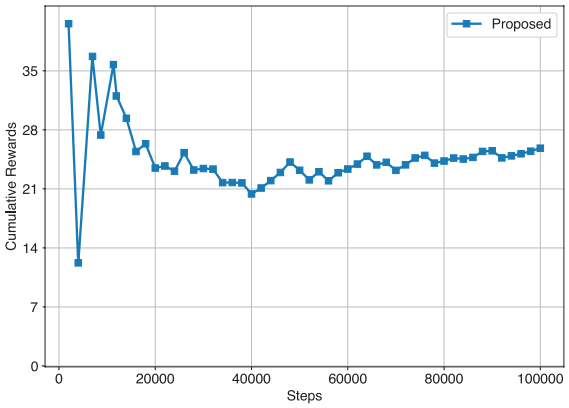

Fig. 2: Convergence of our proposed PPO-based STAR-RIS-assisted MEC system.

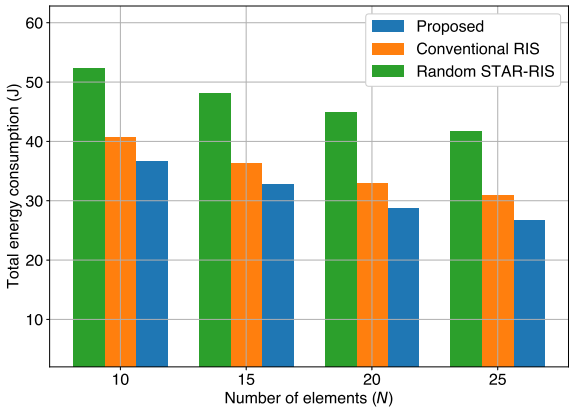

Fig. 4: Performance comparison of total energy consumption under different number of elements

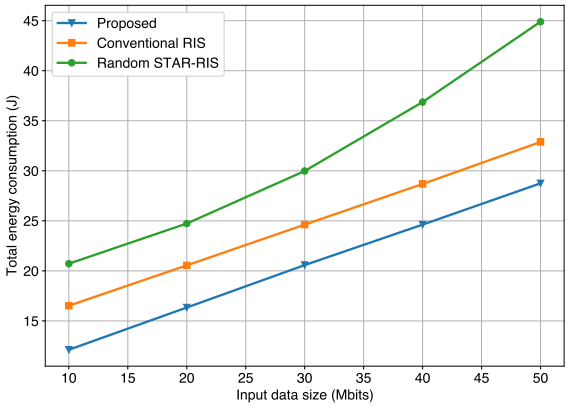

Fig. 3: Performance comparison of total energy consumption under different input data size.

demonstrated stability, sample efficiency, and adaptability in addressing such challenging optimization problems. To prove the efficacy of our proposed algorithm, we have conducted a comprehensive numerical analysis. Based on the performance findings, it can be observed that our proposed method has greater effectiveness compared to several benchmark schemes.

REFERENCES

[1] H. Li, H. Xu, C. Zhou, X. Lü, and Z. Han, "Joint optimization strategy of computation offloading and resource allocation in multi-access edge computing environment," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 10 214–10 226, Jun. 2020.

[2] Y. K. Tun, Y. M. Park, N. H. Tran, W. Saad, S. R. Pandey, and C. S. Hong, "Energy-efficient resource management in UAV-assisted mobile edge computing," *IEEE Communications Letters*, vol. 25, no. 1, pp. 249–253, Sep. 2020.

[3] N. N. Ei, M. Alsenwi, Y. K. Tun, Z. Han, and C. S. Hong, "Energy-efficient resource allocation in multi-uav-assisted two-stage edge computing for beyond 5G networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 16 421–16 432, Feb. 2022.

[4] Y. K. Tun, K. Kim, P. S. Aung, M. Alsenwi, and C. S. Hong, "An efficient resource sharing model for multi-UAV-assisted wireless networks," in *Proc. IEEE Asia-Pacific Network Operations and Management Symposium (APNOMS)*, Virtual Platform, Sep. 2021, pp. 390–393.

[5] P. S. Aung, Y. K. Tun, Z. Han, and C. S. Hong, "Energy-efficiency maximization of multiple RISs-enabled communication networks by deep reinforcement learning," in *Proc. IEEE International Conference on Communications (ICC)*, Seoul, South Korea, 2022.

[6] P. S. Aung, Y. M. Park, Y. K. Tun, Z. Han, and C. S. Hong, "Energy-efficient communication networks via multiple aerial reconfigurable intelligent surfaces: Drl and optimization approach," *IEEE Transactions on Vehicular Technology*, Oct. 2023.

[7] H. Mei, K. Yang, J. Shen, and Q. Liu, "Joint trajectory-task-cache optimization with phase-shift design of RIS-assisted UAV for MEC," *IEEE Wireless Communications Letters*, vol. 10, no. 7, pp. 1586–1590, Apr. 2021.

[8] Q. Zhang, Y. Wang, H. Li, S. Hou, and Z. Song, "Resource allocation for energy efficient STAR-RIS aided MEC systems," *IEEE Wireless Communications Letters*, vol. 12, no. 4, pp. 610–614, Jan. 2023.

[9] P. S. Aung, Y. K. Tun, N. N. Ei, and C. S. Hong, "Energy-efficient offloading and user association in UAV-assisted vehicular ad hoc network," in *Proc. IEEE Asia-Pacific Network Operations and Management Symposium (APNOMS)*, Virtual Platform, Sep. 2020.

[10] Q. Hu, Y. Cai, G. Yu, Z. Qin, M. Zhao, and G. Y. Li, "Joint offloading and trajectory design for UAV-enabled mobile edge computing systems," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1879–1892, Oct. 2018.

[11] P. S. Aung, L. X. Nguyen, Y. K. Tun, Z. Han, and C. S. Hong, "Deep reinforcement learning based joint spectrum allocation and configuration design for star-ris-assisted V2X communications," *IEEE Internet of Things Journal*, Nov. 2023.

[12] Y. K. Tun, N. H. Tran, D. T. Ngo, S. R. Pandey, Z. Han, and C. S. Hong, "Wireless network slicing: Generalized kelly mechanism-based resource allocation," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 8, pp. 1794–1807, Jul. 2019.

[13] P. S. Aung, L. X. Nguyen, Y. K. Tun, Z. Han, and C. S. Hong, "Deep reinforcement learning based spectral efficiency maximization in STAR-RIS-assisted indoor outdoor communication," in *Proc. IEEE/IFIP Network Operations and Management Symposium (NOMS)*, Florida, FL, May. 2023.

[14] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[15] H.-K. Lim, J.-B. Kim, J.-S. Heo, and Y.-H. Han, "Federated reinforcement learning for training control policies on multiple iot devices," *Sensors*, vol. 20, no. 5, p. 1359, Feb. 2020.

[16] A. Juliani, V.-P. Berges, E. Teng, A. Cohen, J. Harper, C. Elion, C. Goy, Y. Gao, H. Henry, M. Mattar *et al.*, "Unity: A general platform for intelligent agents," *arXiv preprint arXiv:1809.02627*, Sep. 2018.