Aerial Reconfigurable Intelligent Surface-enabled SAGIN with LSTM-enhanced DRL Model

Sasinda C. Prabhashana*, Dang Van Huynh*, Keshav Singh† Hans-Jürgen Zepernick§
Octavia A. Dobre*, Hyundong Shin¶, Trung Q. Duong*‡
* Memorial University of Newfoundland, Canada, e-mails:{cwelhengodag, vdhuynh, odobre,tduong}@mun.ca
† National Sun Yat-sen University, Taiwan, e-mail: keshav.singh@mail.nsysu.edu.tw
‡ Queen's University Belfast, UK
§ Blekinge Institute of Technology, Sweden, e-mail: hans-jurgen.zepernick@bth.se
¶ Kyung Hee University, South Korea, e-mail: hshin@khu.ac.kr

***Abstract*—This paper introduces a network architecture that integrates the space-air-ground integrated network with mobile edge computing (MEC) and orbital edge computing to advance sixth-generation (6G) communication systems. The proposed system employs unmanned aerial vehicles (UAVs) equipped with reconfigurable intelligent surfaces and satellite-based MEC to optimize resource management in complex, dynamic environments. By efficiently managing resources such as bandwidth and computational power at both base stations and low Earth orbit satellites, while making offloading decisions, the system aims to minimize utility costs while meeting stringent performance requirements. We utilize a long short-term memory (LSTM)-enhanced deep deterministic policy gradient (DDPG) algorithm to solve the formulated nonlinear programming problem, enabling dynamic and adaptive resource management. The LSTM-enhanced DDPG improves convergence speed by 44.44% compared to conventional DDPG, significantly enhancing cost efficiency. Simulation results validate the robustness of the proposed method against state-of-the-art approaches.**

I. Introduction

The deployment of sixth-generation (6G) networks by 2030 promises to transform global connectivity through comprehensive coverage, improved spectral efficiency, faster data transmission rates, and lower energy consumption and latency [1]. One of the key advancements in 6G is the incorporation of artificial intelligence, which facilitates more intelligent and efficient management of the vast amounts of data and devices present in communication networks [2]. Furthermore, 6G aims to create a seamless integration of satellites and unmanned aerial vehicles (UAVs) within the framework of space-air-ground integrated networks (SAGIN) to ensure fully optimized coverage. In contrast, the current fifth-generation (5G) networks only provide coverage for a small percentage of the world's land area and an even lesser fraction of the Earth's surface, highlighting substantial limitations [3], [4]. To address these obstacles, both innovative and existing technologies need to be enhanced to accommodate the increasing demands.

One notable advancement in communication technology is mobile edge computing (MEC), which provides significant computational resources at the network edge, close to end users. This proximity helps minimize energy consumption in mobile devices, extend battery life, and maintain low latency by offloading computationally intensive tasks to high-performance edge servers [5]. Moreover, SAGIN has led to a paradigm shift in edge-computing-enabled communication services. Terrestrial edge computing is transitioning to non-terrestrial and orbital-edge computing (OEC), finding widespread applications in remote areas [6]. These integrated networks provide ubiquitous connectivity, supporting diverse services such as remote area monitoring, high-speed internet access, and disaster relief, while operating independently [7].

In scenarios involving natural disasters, when terrestrial communication infrastructure may be compromised, unmanned aerial vehicles (UAVs) can facilitate reconnections between users and the nearest available communication systems [8]. Their altitude allows for line-of-sight (LoS) communication with ground base stations, effectively mitigating challenges posed by shadowing and signal obstruction. Additionally, the agility of UAVs enables them to adjust their positions in real time to accommodate fluctuating communication requirements, serving as aerial relays between senders and receivers [9], [10]. Furthermore, reconfigurable intelligent surfaces (RISs) are poised to revolutionize future communication technologies. These arrays consist of adjustable elements that can precisely modify signal phases, thereby enhancing overall communication efficiency [11]. However, many existing implementations of RISs are fixed to locations like walls or rooftops, which can create challenges when obstacles obstruct these surfaces, resulting in diminished system performance [12], [13]. By integrating RISs with UAV technology, it is possible to significantly enhance communication system performance, ensuring a more secure and reliable exchange of information [14].

To optimize resource usage in mobile edge computing (MEC) networks, several methods have been proposed, such as binary task offloading decisions, which reduce edge server idle time and ensure timely responses [15], [16]. However, as user numbers grow and system parameters change rapidly, conventional techniques struggle with efficient decision-making [17]. Deep reinforcement learning (DRL) has shown promise in optimizing real-time decision-making in dynamic environments. DRL agents can efficiently handle complex challenges without prior system knowledge, and recent studies have applied DRL to MEC task offloading, significantly reducing system delays and energy consumption [3], [11], [17]. Integrating DRL with existing technologies enhances resource allocation and addresses growing network demands, warranting further exploration for real-time applications.

In this paper, we present an optimization framework for joint resource allocation, task offloading, and bandwidth management in a MEC-aided SAGIN architecture. Our approach introduces a nonlinear programming problem, aimed at minimizing the total system cost. The formulated optimization problem addresses the dynamic allocation of computational resources, task offloading ratios, and bandwidth distribution across BS. This solution effectively adapts to varying network conditions and user demands, ensuring efficient resource utilization while maintaining low latency. Simulation results demonstrate that the proposed framework significantly enhances performance and cost efficiency, validating its potential for real-world MEC-aided SAGIN applications.

II. System Model and Problem Formulation

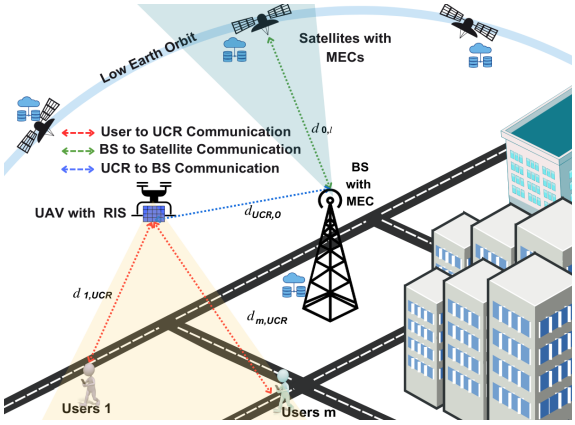

Fig. 1: MEC-aided space-air-ground integrated network.

In this paper, we consider the system architecture for a task offloading strategy aimed at accommodating the resource allocation requirements of end-users through SAGIN, as illustrated in Fig. 1. This model includes a set of M users denoted by $\mathcal{M} = \{1, \ldots, m, \ldots, M\}$, which are registered with a base station (BS). To overcome the limitations posed by non-line-of-sight (NLoS) communication, these users utilize an unmanned aerial vehicle (UAV) equipped with a passive reflective intelligent surface (RIS) with N passive reflecting elements for facilitating make a LoS signal reflection towards the BS. The process of signal reflection is mathematically expressed through the diagonal matrix $\boldsymbol{\phi}$, where $\boldsymbol{\phi} = diag(e^{j\phi_1}, e^{j\phi_2}, \ldots, e^{j\phi_N})$ represents the phase shifts induced by each reflecting element. The BS, equipped with K antennas, incorporates a MEC node to enhance edge computing services for users. Despite these provisions, the BS faces significant challenges in managing the overflow in task requests during peak times, primarily due to the rigorous latency demands of users. To mitigate these challenges, the BS hires $\mathcal{L} = \{1, \ldots, l, \ldots, L\}$ LEO satellites which are in the same circular orbit. Each satellite consists of single antenna and capable of delivering MEC services and guaranteeing the delivery of seamless and dependable services. Moreover, we assume that whenever offloading occurs, a LEO satellite is always in the coverage area and the total coverage time is sufficient to handle and execute the offloaded tasks.

A. Channel Modeling

1) User to BS via UAV-Carried RIS: In this study, we quantify the channel vector of the link between the m-th user and the UCR as $\mathbf{h}_{m,ucr}(t) \in \mathbb{C}^{N\times 1}$ and the channel matrix between the UCR and the BS as $\mathbf{H}_{ucr,0}(t) \in \mathbb{C}^{K\times N}$. We utilize the Rician fading model along with large-scale path loss for channel behavior analysis. Given the dynamic nature of UCR, the effects of the NLoS components are negligible. Consequently, this allows for a simplified expression of the channel gain vectors. Therefore, at time t, $\mathbf{h}_{m,ucr}(t)$ and $\mathbf{H}_{ucr,0}(t)$ can be expressed as in [8].

$$\mathbf{h}_{m,ucr}(t) = \sqrt{\epsilon_0} d_{m,ucr}^{-\delta^{(1)}}(t) \left(\Psi_1^{\text{LoS}} \mathbf{h}_{m,ucr}^{\text{LoS}}(t)\right), \tag{1}$$

$$\mathbf{H}_{ucr,0}(t) = \sqrt{\epsilon_0} d_{ucr,0}^{-\delta^{(1)}}(t) \left(\Psi_1^{\text{LoS}} \mathbf{H}_{ucr,0}^{\text{LoS}}(t)\right), \tag{2}$$

where ϵ_0 represents the path loss at the reference distance. The terms $d_{m,ucr}(t)$ and $d_{ucr,0}(t)$ denote the distance between the m^{th} user and the UCR, and the distance between the UCR and the BS, respectively. The path loss exponent is given by $\delta^{(1)}$, and $\Psi_1^{\text{LoS}} = \sqrt{\frac{\beta 1}{\beta_1+1}}$, where β_1 is the Rician fading factor. At time t, $\mathbf{h}_{m,u}^{LoS}(t) \in \mathbb{C}^{N\times 1}$ is calculated as $\mathbf{h}_{m,ucr}^{LoS}(t) = \left[1, e^{-j\frac{2\pi}{\lambda} d_u \cos(\phi_{\text{AoA}}(t))}, \ldots, e^{-j\frac{2\pi}{\lambda}(N-1) d_u \cos(\phi_{\text{AoA}}(t))}\right]^T$, where λ is the wavelength of the transmission signal, d_u is the uniform spacing between the RIS elements, and $\phi_{\text{AoA}}(t)$ is the angle of arrival (AoA). Furthermore, $\mathbf{H}_{u,bs}^{LoS}(t) \in \mathbb{C}^{K\times N}$ is given by $\mathbf{H}_{u,bs}^{LoS}(t) = \mathbf{a}_{bs}(\phi_{AoD}(t))\mathbf{a}_u^H(\phi_{AoD}(t))$. The steering vector for the BS, $\mathbf{a}_{bs}(\phi_{AoD}(t)) \in \mathbb{C}^{K\times 1}$, is calculated as: $\mathbf{a}_{bs}(\phi_{AoD}(t)) = \left[1, e^{-j\frac{2\pi}{\lambda} d_{bs} \cos(\phi_{AoD}(t))}, \ldots, e^{-j\frac{2\pi}{\lambda}(K-1) d_{bs} \cos(\phi_{AoD}(t))}\right]^T$, where d_{bs} is the spacing between the BS antennas and $\phi_{AoD}(t)$ is the angle of departure (AoD). Similarly, the steering vector for the UCR, $\mathbf{a}_u(\phi_{AoD}(t)) \in \mathbb{C}^{N\times 1}$, represents the phase shifts introduced by the N elements of the RIS as the signal is reflected towards the BS. It is calculated as: $\mathbf{a}_u(\phi_{AoD}(t)) = \left[1, e^{-j\frac{2\pi}{\lambda} d_u \cos(\phi_{AoD}(t))}, \ldots, e^{-j\frac{2\pi}{\lambda}(N-1) d_u \cos(\phi_{AoD}(t))}\right]^T$ [11].

2) BS to LEO Satellite: The link between the BS and a LEO satellite is modeled as a ground-to-air channel, where we consider free space path loss as the path loss model. Therefore, the channel vector $\mathbf{h}_{k,l}(t) \in \mathbb{C}^{1\times K}$ between the k-th antenna and the l-th LEO satellite can be formulated as follows [10]:

$$\mathbf{h}_{k,l}(t) = \left(\frac{4\pi f_c d_{0,l}(t)}{c}\right)^{\frac{-\delta^{(2)}}{2}} \left(\Psi_2^{\text{LoS}} \mathbf{h}_{k,l}^{\text{LoS}}(t) + \Psi_2^{\text{NLoS}} \mathbf{h}_{k,l}^{\text{NLoS}}(t)\right). \tag{3}$$

In (3), the carrier frequency of the transmission signal is denoted by f_c, and c denotes the speed of light. The path loss exponent is given by $\delta^{(2)}$. The terms Ψ_2^{LoS} and Ψ_2^{NLoS} are defined as $\sqrt{\frac{\beta_2}{\beta_2+1}}$ and $\sqrt{\frac{1}{\beta_2+1}}$, respectively, where β_2 is the Rician factor for this link. The vectors $\mathbf{h}_{k,l}^{\text{LoS}}(t)$ and $\mathbf{h}_{k,l}^{\text{NLoS}}(t)$ represent the LoS and NLoS components, respectively [11].

The distance $d_{0,l}(t)$ between the BS and the l-th satellite varies relative to the BS and can be calculated as [6]

$$d_{0,l}(t) = \sqrt{R^2 + (R+r)^2 - 2R(R+r)\cos(\mu(t))}, \quad (4)$$

where R represents the Earth's radius and $r(t)$ denotes the height from the base station (BS) to the LEO satellite, $\mu(t)$ is the geocentric angle, which can be formulated as $\mu(t) = \cos^{-1}\left(\frac{R}{R+r(t)}\cos\alpha\right) - \alpha$ [6]. Here, α is the elevation angle between the BS and the l-th LEO satellite.

B. Communication Model

Users are enabled to offload their computationally heavy tasks to the base station through UCR. At the BS, the instantaneous signal-to-interference-plus-noise ratio (SINR) for m-th user can be expressed as [8]

$$\Gamma_m^{\text{bs}}(t) = \frac{p_m^u \left|\mathbf{H}_{ucr,0}(t)\boldsymbol{\phi}(t)\mathbf{h}_{m,ucr}(t)\right|^2}{\sum_{j=1,j\neq m}^{M} p_j^u \left|\mathbf{H}_{ucr,0}(t)\boldsymbol{\phi}(t)\mathbf{h}_{j,ucr}(t)\right|^2 + z^2(t)}, \quad (5)$$

where p_m^{u} represents the total transmit power, and $z(t)$ is the instantaneous noise power characterized by the Gaussian complex normal distribution $\sim \mathcal{CN}\left(0, \sigma^2\right)$. Therefore, the achievable data rate of the m-th user can be calculated as [8]

$$R_m^{\text{bs}}(t) = b(t)B_{w1}\log_2\left(1 + \Gamma_m^{\text{bs}}(t)\right), \quad (6)$$

where $b_m(t) \in [0,1]$ is the allocated bandwidth coefficient of the m-th user, and B_{w1} is the total bandwidth. We assume that for offloading tasks to a LEO satellite, all K antennas jointly transmit the task. This approach is necessary due to the long distance to the LEO satellite, which requires more power to transmit the tasks. Therefore, the signal-to-noise ratio (SNR) at the l-th LEO satellite for an offloaded task of the m-th user can be formulated as [12]

$$\Gamma^{\text{l}}(t) = \frac{p_0|\sum_{k=1}^{K}\mathbf{h}_{k,l}(t)|^2}{B_{w1}N_0}, \quad (7)$$

where p_0 is the total transmit power of the BS, and N_0 is the single-sided noise spectral density. Therefore, the data rate at the l-th LEO satellite for an offloaded task of the m-th user can be expressed as [12]

$$R_m^{\text{l}}(t) = B_{w1}\log_2\left(1 + \Gamma^{\text{l}}(t)\right), \quad (8)$$

C. Task Offloading Model

We propose the following task offloading model to handle task overflow at the BS during peak times. Let the task from the m-th user be denoted as a 3-tuple $x_m = \{g_m, p_m, T_m^{\max}\}$, where g_m is the size of the task in bits, p_m represents the computational requirement, and $T_m^{\max}$ is the maximum threshold latency. The transmission delay $T_{m,tx}^{\text{bs}}$ and transmission energy $E_{m,tx}^{bs}$ of the m-th user can be formulated as [8]

$$T_{m,tx}^{\text{bs}} = \frac{g_m}{R_m^{\text{bs}}(t)}, \quad E_{m,tx}^{\text{bs}}(t) = p_m^u T_{m,tx}^{\text{bs}}, \quad (9)$$

Therefore, the total transmission utility cost $U_{m,tx}^{bs}$ of the m-th user can be expressed as [16]

$$U_{m,tx}^{\text{bs}}(t) = \psi(t)T_{m,tx}^{\text{bs}} + (1-\psi(t))E_{m,tx}^{\text{bs}}(t), \quad (10)$$

where $\psi(t) \in [0,1]$ is the weighting factor between delay and energy. When a task arrives at the BS, the processing time $T_{m,pr}^{\text{bs}}$ at the BS can be denoted as [12]

$$T_{m,pr}^{\text{bs}} = \frac{p_m}{\eta_m^{(1)}(t)F^{\text{bs}}}, \quad (11)$$

where $\eta_m^{(1)}(t) \in [0,1]$ is the allocated computational power coefficient at the BS for the m-th user's task, and F_{bs} is the total computational power at the BS. Moreover, the energy consumption $E_{m,pr}^{\text{bs}}$ at the BS for the m-th user's task can be formulated as [12]

$$E_{m,pr}^{\text{bs}}(t) = k^{\text{bs}}p_m(\eta_m^{(1)}(t)F^{\text{bs}})^2, \quad (12)$$

where k^{bs} is the energy coefficient of the BS processor, which depends on the capacitance of the integrated chip. Consequently, the total utility cost $U_{m,pr}^{\text{bs}}$ for processing the m-th user's task at the BS is expressed as

$$U_{m,pr}^{\text{bs}}(t) = \psi(t)T_{m,pr}^{\text{bs}} + (1-\psi(t))E_{m,pr}^{\text{bs}}(t). \quad (13)$$

When offloading the entire tasks to the l-th LEO satellite without processing at the BS, the total transmission time $T_{m,tx}^{textl}$ and energy consumption $E_{m,tx}^{\text{l}}$ of the m-th user's task can be denoted as

$$T_{m,tx}^{\text{l}} = \frac{g_m}{R_m^l(t)}, E_{m,tx}^{\text{l}}(t) = p_0 T_{m,tx}^{\text{l}}. \quad (14)$$

Therefore, total utility cost $U_{m,tx}^{\text{l}}$ for offloading the m-th user's task can be expressed as

$$U_{m,tx}^{\text{l}}(t) = \psi(t)T_{m,tx}^{\text{l}} + (1-\psi(t))E_{m,tx}^{\text{l}}(t). \quad (15)$$

When the task from the m-th user arrives at the l-th LEO satellite for processing, the total processing time $T_{m,pr}^{l}$ at the l-th LEO satellite can be expressed as

$$T_{m,pr}^{\text{l}} = \frac{p_m}{\eta_m^{(2)}(t)F^{\text{l}}}, \quad (16)$$

where $\eta_m^{(2)}(t) \in [0,1]$ is the allocated computational power coefficient at l-th LEO satellite for the m-th user's task. F^{l} is the total computation power of the l-th LEO satellite. We consider that all the L satellites have the same computational powers. Moreover, the energy consumption for the execution of the m-th user's task at the l-th satellite can be expressed as

$$E_{m,pr}^{\text{l}}(t) = k^{\text{l}}p_m(\eta_m^{(2)}(t)F^{\text{l}})^2, \quad (17)$$

where k^{l} is the energy coefficient at the l-th LEO satellite processor, which depends on the capacitance of the integrated chip. We consider k^{l} to be the same for the processors at each LEO satellite. Consequently, the total utility cost for processing the m-th user's task at the l-th satellite can be expressed as

$$U_{m,pr}^{\text{l}}(t) = \psi(t)T_{m,pr}^{\text{l}} + (1-\psi(t))E_{m,pr}^{\text{l}}(t). \quad (18)$$

The offloading decision from the BS depends on the total utility cost for each user's task. Therefore, the total utility cost for the m-th user's task can be calculated as

$$\begin{aligned} U_m^{\text{tot}}(t) = {} & \psi(t)T_{m,tx}^{\text{bs}} + (1-\psi(t))E_{m,tx}^{\text{bs}}(t) \\ & + \chi(t)\left(\psi(t)T_{m,pr}^{bs} + (1-\psi(t))E_{m,pr}^{\text{bs}}(t)\right) \\ & + (1-\chi(t))\left(\psi(t)T_{m,tx}^{\text{l}} + (1-\psi(t))E_{m,tx}^{\text{l}}(t)\right. \\ & \left. +\psi(t)T_{m,pr}^{\text{l}} + (1-\psi(t))E_{m,pr}^{\text{l}}(t)\right). \end{aligned} \tag{19}$$

where $\chi(t) \in [0,1]$ is the task offloading fraction determined by the BS based on the utility cost U_m^{tot} of each user's task. Moreover, the total delay for the m-th user's task can be expressed as

$$T_m^{\text{tot}} = T_{m,tx}^{\text{bs}} + \chi(t)T_{m,pr}^{\text{bs}} + (1-\chi(t))\left(T_{m,tx}^{\text{l}} + T_{m,pr}^{\text{l}}\right). \tag{20}$$

D. Problem Formulation

In this paper, we aim to minimize the total utility cost for all $\mathcal{M}$ users during task offloading, which is expressed as

$$\begin{aligned} \Omega(\mathbf{b}, \boldsymbol{\chi}, \boldsymbol{\eta}) = \sum_{m=1}^{M} & U_{m,tx}^{\text{bs}}(t) + \chi_m(t)\left(U_{m,pr}^{\text{bs}}(t)\right) \\ & + (1-\chi_m(t))\left(U_{m,tx}^{\text{l}}(t) + U_{m,pr}^{\text{l}}(t)\right), \end{aligned} \tag{21}$$

where $\mathbf{b} \triangleq \{b_m(t)\}_{\forall m}, \boldsymbol{\chi} \triangleq \{\chi_m(t)\}_{\forall m}, \boldsymbol{\eta} \triangleq \{\eta_m^{(1)}(t), \eta^{(2)}(t)\}_{\forall m}$ are the optimization variables.

Then, the optimization problem is formulated as follows.

$$\begin{aligned} \textbf{\textit{(P1):}} \min_{\mathbf{b},\boldsymbol{\chi},\boldsymbol{\eta}} \quad & \Omega(\mathbf{b}, \boldsymbol{\chi}, \boldsymbol{\eta}), && \text{(22a)} \\ \text{s.t.} \quad & 0 \le b(t) \le 1, \forall m, && \text{(22b)} \\ & 0 \le \chi(t) \le 1, \forall m, && \text{(22c)} \\ & 0 \le \eta^{(1)}(t) \le 1, \forall m, && \text{(22d)} \\ & 0 \le \psi(t) \le 1, \forall m, && \text{(22e)} \\ & 0 \le \eta^{(2)}(t) \le 1, && \text{(22f)} \\ & T_m^{\text{tot}} \le T_m^{\max}, \forall m, && \text{(22g)} \\ & F^{\text{bs}} \ge \eta^{(1)}(t)F^{\text{bs}}, F^{l} \ge \eta^{(2)}(t)F^{\text{l}}, && \text{(22h)} \end{aligned}$$

As outlined in (22), constraint (22b) ensures that the bandwidth allocation coefficient for each user $b(t)$ lies between 0 and 1 for all $m \in \mathcal{M}$. Constraint (22c) requires the offloading fraction $\chi(t)$ to also be between 0 and 1 for all $m \in \mathcal{M}$. Constraints (22d) and (22f) mandate that the computation power allocation coefficients $\eta^{(1)}(t)$ at the BS and $\eta^{(2)}(t)$ at the satellite, respectively, must be between 0 and 1 for all $m \in \mathcal{M}$. Moreover, constraint (22e) sets the weighting factor $\psi(t)$ between delay and energy to be between 0 and 1 for all $m \in \mathcal{M}$. Furthermore, constraint (22g) ensures that the delay for each user's task does not exceed the maximum tolerable delay. Constraint (22h) ensures that the computational resources at the BS and the l-th LEO satellite meet the required allocation: $F^{\text{bs}} \ge \eta^{(1)}(t)F^{\text{bs}}$ and $F^{\text{l}} \ge \eta^{(2)}(t)F^{\text{l}}$.

III. Deep Reinforcement Learning Based Solution

The formulated problem in (22a) is computationally complex and intractable with no feasible solution. As the network devices scale, using the traditional optimization methods are technically impossible. We propose a novel LSTM-enhanced DDPG algorithm to enhance the resource allocation in dynamic environment.

A. MDP Formulation

To apply the LSTM-enhanced DDPG framework for solving (22), we first need to formulate it as a Markov decision process (MDP), characterized by a 3-tuple $\{\mathcal{S}, \mathcal{A}, \mathcal{R}\}$. $\mathcal{S}$ indicates state space, $\mathcal{A}$ is the action space and $\mathcal{R}$ is the reward function.

1) State Space: The state space $s(t)$ represents the environment's information. In this scenario, the state $s(t)$ includes individual utility costs, defined as $s(t) = \{U_1^{\text{tot}}, \ldots, U_m^{tot}, \ldots U_M^{\text{tot}}\}$, where each U_m^{tot} captures the cumulative utility cost of user m over time.

2) Action Space: The action space $a(t)$ consists of key decision variables that the agent controls to optimize system performance. This action space is represented as $a(t) = \{b(t), \chi(t), \eta^{(1)}(t), \eta^{(2)}(t)\}$.

3) Reward: The reward function can be formulated as the inverse of the system's total utility cost, as expressed in the following equation:

$$\begin{aligned} r(t) = \Bigg(\sum_{m=1}^{M} & \left(U_{m,tx}^{\text{bs}} + \chi(t)U_{m,pr}^{\text{bs}} \right. \\ & \left. + (1-\chi(t))\left(U_{m,tx}^{\text{l}} + U_{m,pr}^{\text{l}}\right)\right)\Bigg)^{-1}. \end{aligned} \tag{23}$$

The utility cost for each user m is adjusted by incorporating a penalty ϵ if any of the constraints C_1 through C_8 are not satisfied. This adjustment is represented by $U_m^{\text{tot}} = U_m^{\text{tot}} + \epsilon$. This formulation ensures that a reduction in the total utility cost of the system, while maintaining adherence to all constraints, results in an increase in the reward $r(t)$.

B. LSTM-enhanced DDPG Algorithm

To enhance the optimization capabilities within the formulated MDP framework, we employ the DDPG algorithm, a state-of-the-art model-free, actor-critic DRL technique. It involves two primary networks: the actor network $\mu(s|\theta^{\mu})$, which generates actions given states, and the critic network $Q(s, a|\theta^{Q})$, which evaluates these actions. The actor network includes an LSTM layer that processes each state individually. The detailed algorithm of the proposed LSTM-enhanced DDPG solution is in Algorithm 1.

IV. Numerical Results and Discussions

A. Simulation Settings

This subsection presents the settings of parameters for the implementation of the proposed solution and simulations. Firstly, the replay buffer of the LSTM-enhanced DDPG algorithm can store up to 100,000 experience transitions and 32 mini-batches at a time. The actor learning rate is set to

0.0001 and the critic learning rate to 0.001. The discount factor γ is 0.99, and the soft update parameter τ is 0.001. The temporal correlated noise adopts the ornstein uhlenbeck process with mean reversion rate is 0.15 and volatility is 0.2. We compare the total rewards over 1,000 testing episodes, with each episode consisting of 100 time slots. Thus, the total reward earned in an episode is the sum of the rewards earned across all 100 time slots.

Algorithm 1 : Proposed LSTM-enhanced DDPG Algorithm for solving (22).

Initialize the environment with its specified parameters.
Initialize the actor network $\mu(s|\theta^{\mu})$ with an LSTM layer and parameters θ^{μ} and the critic network $Q(s,a|\theta^{Q})$ with parameters θ^{Q}.
Initialize the target networks μ' and Q' with $\theta^{\mu'} \leftarrow \theta^{\mu}$ and $\theta^{Q'} \leftarrow \theta^{Q}$.
Set up a replay buffer $\mathcal{R}$.
for each episode **do**
  for $t = 1, 2, \ldots, T$ **do**
    Process the state $s(t)$ through the LSTM layer in the actor network to generate the action $a(t)$;
    Generate an action with added noise: $a(t) = \mu(s(t)|\theta^{\mu}) + Z(t)$;
    Execute the action $a(t)$, then receive reward $r(t)$ and the next state $s(t+1)$;
    Store $\{s(t), a(t), r(t), s(t+1)\}$ in the replay buffer;
    Randomly sample a batch S from the replay buffer;
    Compute target values using target networks and store in y_i using equation:
    $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$;
    Update the critic network parameters θ^{Q} by minimizing the loss L using equation:
    $L = \frac{1}{S}\sum_i \left(y_i - Q(s_i, a_i|\theta^{Q})\right)^2$;
    Compute policy gradients and update the actor network parameters θ^{μ} using gradient $\nabla_{\theta^{\mu}} J$ from equation:
    $\nabla_{\theta^{\mu}} J = \frac{1}{S}\sum_i \left(\nabla_a Q(s_i, a_i|\theta^{Q})\big|_{a_i=\mu(s_i|\theta^{\mu})} \nabla_{\theta^{\mu}} \mu(s_i|\theta^{\mu})\right)$
    Softly update the target networks using the soft update rule with coefficient τ according to equations
    $\theta^{\mu'} \leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'}$ and
    $\theta^{Q'} \leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'}$;
  end for
end for

Moreover, for the other network parameters, the number of users are $M = 10$, the number of RIS elements are $N = 16$, and the number of antennas at the base station are $K = 8$. The Earth's radius is 6371 km, and the distance to LEO satellites is $r = 500$ km. The distance from the m^{th} user to the UAV is $d_{m,ucr} = 100$ m, and the distance from the UAV to the base station is $d_{ucr,0} = 100$ m. The path loss exponent values are $\delta^{(1)} = 3.65 - 3.75$ and $\delta^{(2)} = 2$. The base station MEC computation power is $F^{bs} = 4$ GHz, while the satellite MEC computation power is $F^{l} = 2$ GHz. The transmission power of each user is $P_m^u = 20$ dBm, with a task size $g_m \in [5, 6]$ Mbits, and task complexity $p_m \in [500, 600]$ Mcycles. The transmit power of the base station is $p_0 = 40$ dBm, and the system bandwidth is $B_{w1} = 50$ MHz. The noise power is -110 dBm/Hz, and the maximum delay for each task is $T_k^{\max} = [5, 6]$ s. The energy coefficients for the BS and satellites are $k^{bs} = 1 \times 10^{-27}$ and $k^{l} = 1 \times 10^{-27}$, respectively. The carrier frequency is $f_c = 2$ GHz, and the balancing factor is $\psi = 0.5$ [16].

B. Numerical Results

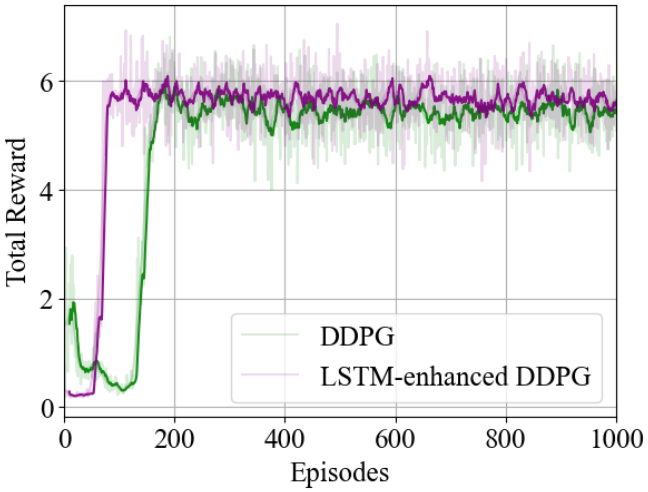

Fig. 2: Convergence performance.

1) Performance of Convergence: The convergence of the LSTM-enhanced DDPG and the standard DDPG algorithms is depicted in Fig. 2. It is evident that the LSTM-enhanced DDPG demonstrates significantly faster convergence compared to the standard DDPG. Initially, both algorithms exhibit low rewards, reflecting the exploration phase. However, the LSTM-enhanced DDPG quickly outperforms the standard DDPG, reaching a stable and higher reward after fewer episodes. The LSTM-enhanced DDPG maintains more consistent rewards throughout the learning process, indicating better overall performance and stability in learning optimal policies over time.

2) Average utility cost for different system bandwidth: : The relationship between system bandwidth and average utility cost is depicted in Fig. 3. It is evident that as bandwidth increases, the utility cost decreases for both DDPG and LSTM-enhanced DDPG. However, the LSTM-enhanced DDPG consistently achieves lower utility costs across all bandwidth levels, particularly under constrained bandwidth conditions.

3) Average utility cost for different task complexities: The impact of task complexity on average utility cost is illustrated in Fig. 4. As task complexity increases, the utility cost rises for both algorithms. However, the LSTM-enhanced DDPG consistently incurs lower utility costs compared to the standard DDPG across all complexity levels. The difference in performance becomes more pronounced with higher task complexities, demonstrating the LSTM-enhanced DDPG's ability to

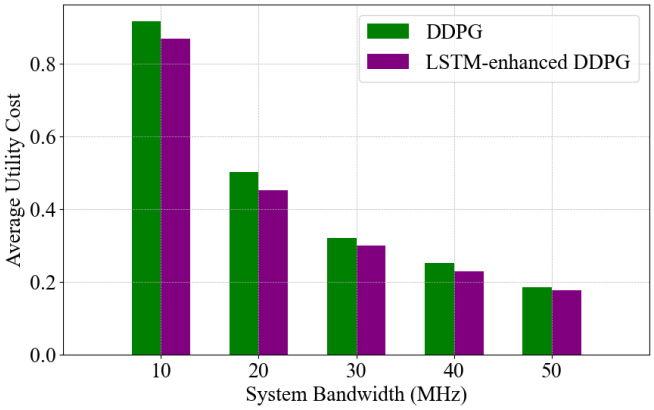

Fig. 3: Average utility cost for different system bandwidth.

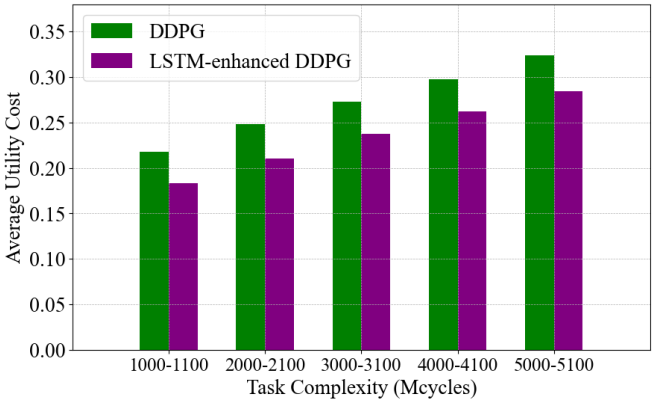

Fig. 4: Average utility cost for different task complexities.

handle more computationally intensive tasks while optimizing resource allocation more effectively. This confirms that the LSTM-enhanced approach is particularly robust in managing complex scenarios, ensuring more efficient use of available resources.

V. Conclusion

This paper introduced a novel architecture for the SAGIN, leveraging MEC with UAV-carried RIS to optimize resource management in dynamic 6G environments. The proposed architecture minimizes total system utility costs through optimal allocation of bandwidth, computational power at BS and satellites, and offloading decisions. We employed an LSTM-enhanced DDPG algorithm to effectively manage the distribution of computational tasks between terrestrial and non-terrestrial components, ensuring efficient utilization of resources while meeting stringent performance requirements. Simulation results demonstrated that the LSTM-enhanced DDPG significantly outperforms conventional method, leading to substantial improvements in system performance and cost efficiency. These findings highlight the potential of the proposed SAGIN architecture in advancing real-world applications of integrated networks.

References

[1] K. Trichias, A. Kaloxylos, and C. Willcock, "6G global landscape: A comparative analysis of 6G targets and technological trends," in *Proc. Joint European Conf. on Net. and Commun. and 6G Summit (EuCNC/6G Summit)*, Gothenburg, Sweden, Jun. 3–6 2024.

[2] S. Yrjölä, M. Matinmikko-Blue, and P. Ahokangas, "Developing 6G visions with stakeholder analysis of 6G ecosystem," in *Proc. Joint European Conf. on Net. and Commun. and 6G Summit (EuCNC/6G Summit)*, Gothenburg, Sweden, Jun. 6–9 2023.

[3] Q. Dong, X. Xu, S. Han, R. Liu, and X. Zhang, "DDPG-based task offloading in satellite-terrestrial collaborative edge computing networks," in *Proc. IEEE Int. Conf. on Commun. Workshops (ICC Workshops)*, Rome, Italy, 28 May – 01 Jun 2023.

[4] T. T. Bui, A. Masaracchia, V. Sharma, O. Dobre, and T. Q. Duong, "Impact of 6g space-air-ground integrated networks on hard-to-reach areas: Tourism, agriculture, education, and indigenous communities," *EAI Endorsed Transactions on Tourism, Technology and Intelligence*, vol. 1, no. 1, pp. 1–8, Sep. 2024.

[5] Y. Qian, J. Xu, S. Zhu, W. Xu, L. Fan, and G. K. Karagiannidis, "Learning to optimize resource assignment for task offloading in mobile edge computing," *IEEE Commun. Lett.*, vol. 26, no. 6, pp. 1303–1307, Jun. 2022.

[6] K. Wei, Q. Tang, J. Guo, M. Zeng, Z. Fei, and Q. Cui, "Resource scheduling and offloading strategy based on LEO satellite edge computing," in *Proc. IEEE 94th Veh. Technol. Conf. (VTC-Fall)*, Norman, OK, USA, Sep. 27–30 2021, pp. 1–6.

[7] A. Umar, S. A. Hassan, and H. Jung, "Computation offloading and resource allocation in NOMA-MEC enabled aerial-terrestrial networks exploiting mmwave capabilities for 6G," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Denver, CO, USA, Jun. 9–13 2024.

[8] G. Sun, L. He, Z. Sun, Q. Wu, S. Liang, J. Li, D. Niyato, and V. C. M. Leung, "Joint task offloading and resource allocation in aerial-terrestrial UAV networks with edge and fog computing for post-disaster rescue," *IEEE Trans. Mobile Comput.*, vol. 23, no. 9, pp. 8582–8594, Sep. 2024.

[9] Y. Gao, F. Lu, P. Wang, W. Lu, Y. Ding, and J. Cao, "Resource optimization of secure data transmission for UAV-relay assisted maritime MEC system," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Rome, Italy, May 28–Jun. 01 2023, pp. 3345–3350.

[10] M.-H. T. Nguyen, T. T. Bui, L. D. Nguyen, E. Garcia-Palacios, H.-J. Zepernick, H. Shin, and T. Q. Duong, "Real-time optimized clustering and caching for 6G satellite-uav-terrestrial networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 3, pp. 3009–3019, Jun. 2024.

[11] K. K. Nguyen, S. R. Khosravirad, D. B. da Costa, L. D. Nguyen, and T. Q. Duong, "Reconfigurable intelligent surface-assisted multi-UAV networks: Efficient resource allocation with deep reinforcement learning," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 3, pp. 358–367, Apr. 2022.

[12] M. Zhang, Z. Su, Q. Xu, Y. Qi, and D. Fang, "Energy-efficient task offloading in UAV-RIS-assisted mobile edge computing with NOMA," in *Proc. IEEE Int. Conf. on Computer Commun. workshops (INFOCOM workshops)*, Vancouver, Canada, May 20–23 2024, pp. 1–6.

[13] H. Yu, H. D. Tuan, A. A. Nasir, T. Q. Duong, and H. V. Poor, "Joint design of reconfigurable intelligent surfaces and transmit beamforming under proper and improper Gaussian signaling," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2589–2603, Nov. 2020.

[14] J. Yuan, G. Chen, M. Wen, D. Wan, and K. Cumanan, "Security-reliability tradeoff in UAV-carried active RIS-assisted cooperative networks," *IEEE Commun. Lett.*, vol. 28, no. 2, pp. 437–441, Feb. 2024.

[15] M. Mukherjee, V. Kumar, A. Lat, M. Guo, R. Matam, and Y. Lv, "Distributed deep learning-based task offloading for UAV-enabled mobile edge computing," *IEEE Commun. Lett.*, vol. 28, no. 2, pp. 437–441, Feb. 2024.

[16] N. Waqar, S. A. Hassan, A. Mahmood, K. Dev, D. T. Do, and M. Gidlund, "Computation offloading and resource allocation in MEC-enabled integrated aerial-terrestrial vehicular networks: A reinforcement learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 21 478–21 490, Nov. 2022.

[17] N. Sharma, A. Ghosh, R. Misra, and S. K. Das, "Deep meta q-learning based multi-task offloading in edge-cloud systems," *IEEE Trans. Mobile Comput.*, vol. 23, no. 4, pp. 2583–2597, Apr. 2024.