Independent Deep Reinforcement Learning for Optimization of RSMA-enabled hybrid RAN Slicing

Haiyan Tu, Paolo Bellavista, *Senior Member, IEEE,* Liqiang Zhao, *Member, IEEE,* Kai Liang *Member, IEEE,* Gan Zheng, *Fellow, IEEE,* Kai-Kit Wong, *Fellow, IEEE*

***Abstract*—RAN slicing has been widely studied for providing ultra-reliability low-latency communication (URLLC), enhanced mobile broadband (eMBB), and massive machine type communication (mMTC) services in 5G. However, the existing RAN slicing networks have not been explored to support the hybrid services, such as massive URLLC (mULC) and ubiquitous eMBB (uMBB) services. In this paper, we propose a novel rate splitting multi-access (RSMA)-enabled RAN slicing system to facilitate the runtime support of mULC and uMBB services. Firstly, three typical slices, i.e., URLLC, eMBB, and mMTC slices are constructed. Then, a multi-connection scheme is proposed by using RSMA technology, i.e., the users can be connected with two typical slices to obtain mULC and uMBB services. Specifically, the transmitted data of each mULC/uMBB user will be split into the common mMTC data and the private URLLC/eMBB data, which will be encoded into the corresponding traffic flows and served by corresponding slices. Next, a system-wide utility optimization problem is proposed to optimize heterogeneous requirements for mULC and uMBB services by joint user grouping, bandwidth allocation, and power control. Finally, a two independent agent DDPG (2IADDPG) algorithm is customized to solve the formulated problem, wherein two independent agents are responsible for independent decision-making. The reported numerical results show that the RSMA scheme outperforms the benchmarks, and in the meanwhile our proposed 2IADDPG algorithm can achieve faster convergence rate compared with the multi-agent DDPG algorithm and other comparison algorithms.**

***Index Terms*—Rate splitting multi-access, hybrid services, independent Q learning, deep deterministic policy gradient.**

I. Introduction

Radio access network (RAN) slicing [1] may be inherited as one of the key enabling technologies to support the various demands. According to specific requirements, RAN slicing can simultaneously construct multiple virtual networks based upon the same physical network infrastructure and resources, to provide different types of customized services. With the rapid development of the next generation wireless communications, the hybrid services [2], [3], such as massive ultra-reliability low-latency communication (mULC) and ubiquitous enhanced mobile broadband (uMBB) services have emerged, which are the type of critical mMTC [4]. For example, in-space cellular backhaul remote connectivity systems request uMBB service, which requests high broadband data rates along with massive connectivity [5], while metaverse streaming [6] is considered as a type of mULC service, which supports massive numbers of mobile users demanding stringent quality of service requirements. In the current RAN slicing system, the mobile network operators (MNOs) usually construct three typical slices to correspondingly support the ultra reliable low latency communication (URLLC), enhanced mobile broadband (eMBB), and massive machine type communication (mMTC) services. However, there are very limited works on RAN slicing systems supporting hybrid services, which may have some technical challenges that need to be further studied. Therefore, we expect to provide hybrid services on the top of the existing slicing system. To the best of our knowledge, this is the first work that investigates slicing for supporting hybrid services.

The resource isolation is always assumed between slices, which means that there is no inter-slice interference. However, as the significant increase of the types of services and the number of users, the orthogonal-based resource allocation policy is not always efficient. Some works considered the inter-slice interference, e.g., Zambianco *et al.* [7] studied a resource allocation slicing policy to enforce inter-slice isolation across mobile virtual network operators (MVNO) by minimizing the inter-slice interference. However, the resource reuse in these approaches is passive and disordered, which poses great challenges to the interference coordination mechanism and makes it difficult to achieve efficient resource management and performance assurance in high-load network environments. Thus, non-orthogonal multi-access policies are starting to demonstrate that they can achieve good and suitable performance results [8]. Among them, rate splitting multi-access (RSMA) [9] has emerged as a highly-reliable and spectrum-efficient multiple access scheme, which divides a user's message into a private part and a common part, where the common parts are jointly encoded into a common flow for decoding by multiple users, while the private parts are independently encoded into private flows decoded by the corresponding users. At the receiver, the common flow is decoded first and

Haiyan Tu is with the School of Information Engineering, East China University of Technology, Nanchang 330013, China. She is also with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an, 710071, China (e-mail: hytu@stu.xidian.edu.cn).

Paolo Bellavista is with the Department of Computer Science and Engineering, University of Bologna, Bologna, 40126, Italy (e-mail: paolo.bellavista@unibo.it).

Liqiang Zhao is with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an, 710071, China (e-mail: lqzhao@mail.xidian.edu.cn).

Kai Liang is with the School of Telecommunications Engineering, Xidian University, Xi'an, 710071, China (e-mail: kliang@xidian.edu.cn). He is also affiliated with Anhui Province Key Laboratory of Cyberspace Security Situation Awareness and Evaluation, Hefei, 230037, China.

Gan Zheng is with the School of Engineering, University of Warwick, Coventry, CV4 7AL, UK (e-mail: gan.zheng@warwick.ac.uk).

Kai-Kit Wong is affiliated with the Department of Electronic and Electrical Engineering, University College London, Torrington Place, WC1E 7JE, United Kingdom (e-mail: kai-kit.wong@ucl.ac.uk). He is also affiliated with Yonsei Frontier Lab, Yonsei University, Seoul, Korea.

next the private flow is decoded with successive interference cancellation (SIC).

In RSMA, the common messages of different users in the same RSMA group are encoded into a common flow by using the same codebook, while the private messages can be encoded into separate private flows with different codebooks. This design inherently offers a certain level of privacy protection, as user-specific data encoded in private messages remains less exposed compared to encoding it in the common flow. Specifically, with this encoding scheme, if any part of the common flow is compromised, it implies that the eavesdropper has obtained the common codebook, effectively exposing the entire common flow to the open network. In contrast, even if one private flow is compromised, the other private flows remain secure due to the use of distinct codebooks. Furthermore, mMTC traffic is characterized by small traffic size, high homogeneity, and high overlapping terminal interests, which is inherently suitable for multicast transmission through unified coding. Combined with the coding mechanism of RSMA, the mMTC traffic of the hybrid services can be transmitted as the common part of RSMA for broadcast transmission. Supported by these, we use rate-splitting (RS) [10] to take advantage of the correlations resulting from the common parts of two hybrid services [11]. Specifically, the data of each user requesting hybrid services may be split into two parts, i.e., the common mMTC data and the private URLLC/eMBB data. Moreover, we design a multi-connection RAN slicing scheme, that is, a mULC or uMBB user will be connected to two slices to meet the hybrid requirements.

Another crucial and ongoing technical issue for RAN system is radio resource control. Due to the limited resources and time-varying nature of wireless channels in RAN domain, the resource allocation and management issue will be a long-term research topic. The traditional algorithms for resource allocation problems were extensively studied in the literature. For example, Zhou *et al.* [12] relied on the Lyapunov optimization to carry out joint virtual resource optimization to maximize the defined utility function of the RAN slicing system. However, the traditional methods usually have low-efficiency and high-computational complexity, the deep reinforcement learning (DRL)-based algorithms [13] have emerged as efficient approaches for solving high-complexity and non-convex problems. Among them, multi-agent DRL (MADRL), which incorporates multi-agent learning, has received an increased amount of attention recently. MADRL approaches can be classified into two major categories: independent action learners (IAL) [14], [15] and joint action learners (JAL) [16]. With the benefits of independent decision-making for each agent in independent reinforcement learning (IQL) algorithms, we design an IQL-based algorithm in this paper.

Thus, we propose an RSMA-enabled RAN slicing scheme for the hybrid services. Firstly, the three typical slices are constructed in the system: URLLC, eMBB, and mMTC slices. Then, based on the existing RAN slicing system, we design a multi-connection scheme for supporting the hybrid services by introducing RSMA. In RSMA, the data of each mULC or uMBB user is split into two parts: common mMTC part and private URLLC/eMBB part. Each part will be served by the corresponding slice, i.e., each user will be connected with two slices. Furthermore, a utility maximization problem is constructed by jointly optimizing user grouping and resource allocation. Finally, a novel two independent agents deep deterministic policy gradient (2IADDPG) algorithm is proposed to solve the problem. Importantly, though only two kinds of hybrid services are considered in this paper, our proposed solution can dynamically adapt to the changing scenarios. The main contributions of this paper are as follows:

- ***RSMA-based Multi-connection RAN Slicing Scheme***: Firstly, three basic RAN slices are constructed upon the same underlying physical network to collaboratively provide the hybrid services, namely URLLC, eMBB, and mMTC slices. Especially, we consider two kinds of hybrid services in this paper, which are the mULC and uMBB services. Then, we propose a multi-connection RAN slicing scheme based on RSMA, i.e., each mULC or uMBB user is connected with multiple basic slices to meet the hybrid requirements. Based on RS, the data of users can be split into two parts (common mMTC data and private URLLC/eMBB data), and each of them is connected with the corresponding slice to obtain services. Naturally, the mULC and uMBB users are grouped into pairs to facilitate the application of RSMA.
- ***Utility Maximization Problem Formulation***: We aim at maximizing the system-wide utility by jointly optimizing user grouping, subchannel allocation, and power control, which is a non-deterministic polynomial-time hard (NP-hard) problem. Specifically, the utility function is calculated as the weighted sum of three performance metrics: throughput, reliability delay, and coverage probability, to meet the multiple requirements of the hybrid services.
- ***2IADDPG algorithm***: Considering the independent relationship between the user grouping, subchannel allocation, and power allocation issues, we propose a novel 2IADDPG algorithm to solve the utility maximization problem. The algorithm obtains the solution of the user grouping and subchannel allocation by the first agent. Then, the user grouping results of the first agent will be input as the partial states of the second agent, and the second agent is responsible for the power control.

Numerical results show that our proposed RSMA scheme outperforms the compared schemes in terms of the coverage probability, throughput, and reliable delay. The proposed algorithm achieves a faster convergence rate, while significantly improve the performance.

The rest of the paper is organized as follows. In the next section, we briefly introduce the related work. In Section III, our RSMA-enabled RAN slicing system model is proposed, and the different RAN slices are described to collaboratively provide hybrid services. In Section IV, we formulate the optimization problem of the RSMA-enabled RAN slicing system. In particular, the targeted optimization problem is formulated to meet the different quality of service (QoS) requirements of the hybrid services. In Section V, we solve the proposed problem by customizing an independent DRL model structure and propose a 2IADDPG algorithm. Section VI is dedicated

to the extensive discussion of the reported simulation results, while the conclusion remarks in Section VII end the paper.

II. Related Work

In this section, we will illustrate the state-of-the-art in RAN slicing and RSMA, as well as the introduction of the DRL-based algorithms.

A. Customized services provided by RAN slicing

RAN slicing has been widely studied to provide customized services, by sharing the same physical infrastructure. Setayesh *et al.* [17] studied the eMBB and URLLC network slices by sharing the same RAN infrastructure, wherein the punctured scheduling method is adopted between the eMBB and URLLC users. Alcaraz *et al.* [18] investigated a model-based reinforcement learning approach to efficiently manage the resource allocation among the eMBB and mMTC slices. In addition, multiple vehicle-to-everything (V2X) slices were constructed to provide customized services for UL/DL Decoupled Cellular V2X networks in [19]. Although there are many works about RAN slicing, nevertheless, they only studied a single type of services and there is a paucity of literature on the hybrid services [20], such as mULC and uMBB services. Zeng *et al.* [21] considered the energy-efficient mULC scenario, which integrates URLLC with massive access, over the cell-free (CF) massive multiple-input–multiple-output (MIMO) system. Zhang *et al.* [22] developed analytical models for CF massive MIMO system to support the new 6G standard traffic services, which is mULC communications. The above works focused on the mULC services, they did not consider the coexistence of mULC and uMBB services. To the best of our knowledge, there are very limited works on RAN slicing to provide hybrid services. Encouraged by these, we develop a cooperative RAN slicing solution for supporting the mULC and uMBB simultaneously.

B. RSMA scheme in wireless network

RSMA is proven to improve energy efficiency, reliability, and delay at a lower computational complexity. Singh *et al.* [23] considered a downlink wireless network consisting of an unmanned aerial vehicle (UAV)-assisted base station (BS), where RSMA is introduced to serve multiple ground users (GUs) simultaneously. Xia *et al.* [24] explored the security-reliability trade-off in RSMA-based beam-forming against eavesdropper collusion, which aimed to maximize the minimum secrecy rate (MSR) while considering user fairness. The authors in [25] proposed two RSMA-based strategies, namely, time partitioning-RSMA (TP-RSMA) and power partitioning-RSMA (PP-RSMA), where PP–RSMA was approved as a powerful physical-layer transmission approach for overloaded cellular internet of things (IoT). Then, Cho *et al.* [26] proposed a cooperative RSMA scheme to increase the coverage for the downlink system in a THz scenario. The work in [27] applied RSMA to an uplink two-user single-input single-output (SISO) multiple access channel communication system to improve the error probability performance, sum-throughput, and the rate region. In addition, some works have studied RSMA-based slicing schemes. Santos *et al.* [28] adopted an RSMA-based radio resource slicing strategy for URLLC uplink transmission, in which the URLLC message is split into two sub-messages. Liu *et al.* [29] studied the RSMA-based slicing scheme, and the results show that RSMA can outperform NOMA counterpart in network slicing, and obtain significant gains over OMA in some regions. With the above advantages, and considering the common parts (mMTC part) of the two hybrid services, RSMA can be beneficial for our model.

C. DRL-based algorithm

DRL [30] has shown great potential in addressing the communication, computing, caching, and control (4Cs) problems. Jiang *et al.* [31] proposed a Q-MIX and proximal policy optimization (PPO) algorithm to solve the long-term optimization problem in the multi-access edge computing (MEC) network slicing system. Azimi *et al.* [32] applied the asynchronous advantage actor-critic (A3C) algorithm to optimize the energy-efficient power allocation problem. DDPG [33], [34] is widely used to cope with continuous-valued control problems and solve the non-convex objective function in a long-term average form. However, the single-agent DRL algorithms may not be able to cope with increasingly complex environments. In this case, multi-agent DRL algorithms [35], [36] have emerged and been applied to effectively solve complex and high-dimension optimization problems. Boateng *et al.* [37] proposed a novel stackelberg multi-agent deep deterministic policy gradient (MADDPG) algorithm for slice creation and autonomous resource allocation. Andreou *et al.* [38] proposed a comprehensive strategy for network slicing design and applied the MADDPG algorithm to the configuration of network slices, and to enhance network efficiency and performance. Furthermore, IQL algorithms may perform on par or better than multi-agent algorithms, even in more challenging environments [39]. Hu *et al.* [40] let the independent agents compute common knowledge information for action selection to mitigate the effects of environment non-stationarity, which outperforms multi-agent common knowledge reinforcement learning. In [41], the authors represent an independent DQN agents-based scheme to support dynamic slice embedding and reconfiguration. The IQL-based algorithms learn strategies based on local observation, allowing each agent to independently generate actions and update the strategies, leading to faster convergence. This motivates our work on the IQL-based optimization method.

III. System Model

We consider an RSMA-enabled downlink RAN slicing system as shown in Fig. 1, where the users randomly request the hybrid mULC and uMBB services. The potential example applications of the hybrid services are extensive. Taking industrial automation as an example, in industrial automation scenarios, industrial robots require massive connectivity (mMTC data) to support the cooperative operation of multiple devices, and rely on millisecond-level ultra-low-latency control

signals (URLLC data) to ensure precise execution of operations simultaneously. In addition, the high-definition cameras carried by robots require high-throughput video transmission (eMBB data) for AI real-time visual inspection to determine product quality and business qualification. In these complex application scenarios, the existence of mULC and uMBB services can efficiently meet the comprehensive requirements of low latency, high reliability, large data throughput, and connection density of the overall system, so we consider the hybrid services in this paper.

In this model, three basic slices are supported in the system, namely the URLLC, eMBB, and mMTC slices. They are referred as slice s, $s \in \{U, E, C\}$. Then, RSMA is introduced in this model, where the data of users will be split into common mMTC data, and private URLLC or eMBB data[1]. From a security perspective, mMTC data (which is mainly related to connectivity and coverage commands) is encoded as the common message, while URLLC and eMBB data (potentially containing user-specific or privacy-sensitive information) are treated as private messages. By encoding sensitive or privacy-critical information into private messages rather than the common flow, can enhance data security and mitigates the risk of exposure, ensuring a more robust and secure transmission. Moreover, mMTC traffic, with small packets and high homogeneity, is ideal for multicast via unified coding in the common part of RSMA. Thus, the mMTC data of all users in the same RSMA group is encoded into a common flow, while the URLLC and eMBB data is encoded into different private flows. Further, the flows are served by the corresponding slices, i.e., the users are in multi-connection mode. Thus, each user will be connected with two slices to obtain the hybrid services. In addition, the users are divided into disjoint groups to use RSMA, the set of which is $\mathcal{G} = \{1, 2, \cdots, G\}$. Importantly, this method is applicable to hybrid services with separable traffic and cannot be generalized to hybrid services with non-separable traffic. In industrial automation applications, due to isolated subsystems within the robots, the different types of traffic are generated by independent hardware modules, thereby enabling traffic classification and splitting.

Without loss of generality, we assume N BSs and K users are randomly distributed, in which K_U users would like to request mULC service and K_M users request uMBB service, and we have $K = K_U + K_M$. To simplify the analysis, we assume $K_U = K_M$ and two heterogeneous users (one uMBB user and one mULC user) are paired in a group to facilitate the use of RSMA. Then we have $G = K/2$. Let $\mathbf{c}(t) = \{c_{k,g}(t), \forall k, g\}$ denote the user grouping indicators, where $c_{k,g}(t) = 1$ means user k is in group g, otherwise not. The maximum available power of each BS is assumed to be P_B. We assume that there are totally T times slots, where each time slot is defined as the time interval $[t, t+1]$, $t \in \{0, 1, 2, \cdots, T-1\}$.

[1]The two kinds of data of each hybrid service can be categorized based on the traffic identification methods. Moreover, RSMA allows different traffic types to be encoded in different ways [8], making it a viable method for traffic differentiation. Then, the identification results can provide theoretical support for data splitting.

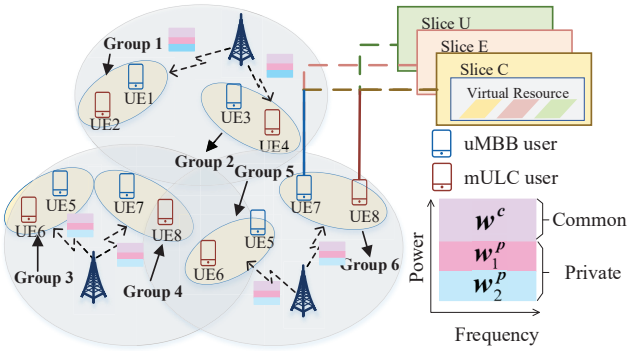

Fig. 1. RSMA-enabled RAN slicing system with hybrid services.

The total bandwidth is set to B, divided into M orthogonal subchannels with the set of $\mathcal{M} = \{1, 2, \cdots, M\}$. Let the binary variables $\mathbf{x}(t) = \{x_g^{b,m}(t), \forall g, m, b\}$ indicate the subchannel allocation variables, where $x_g^{b,m}(t) = 1$ represents group g is allocated with subchannel m of BS b at time slot t, otherwise not.

Then, the received signal to interference plus noise ratio (SINR) of user k on slice C in group g decoding the common mMTC flow with subchannel m at time slot t is expressed as:

$$\gamma_{k,g}^{c,m}(t) = \frac{P_g^{c,m}(t)|h_k^m(t)|^2}{I_{k,g}^{c,m}(t) + N_0}, \tag{1}$$

where $P_g^{c,m}(t)$ is the power allocated to the common flow of group g on slice C with subchannel m, while $h_k^m(t)$ represents the channel gain of user k in group g with subchannel m at time slot t and N_0 is the noise power. In addition, $I_{k,g}^{c,m}(t)$ is the interference of user k on slice C in group g with subchannel m, which is defined as

$$\begin{aligned} I_{k,g}^{c,m}(t) = &\sum_{p\in\{u,e\}} P_{k,g}^{p,m}(t)|h_k^m(t)|^2 + \\ &\sum_{g'\in\mathcal{G}\setminus g}\sum_{i=1}^{K} x_{g'}^{b,m}(t) P_{i,g'}^m(t)|h_k^m(t)|^2, \end{aligned} \tag{2}$$

where $P_{i,g'}^m(t) = \sum_{j\in\{c,u,e\}} P_{i,g'}^{j,m}(t)$, and $P_{k,g}^{p,m}(t)$ is the power allocation of private flow on slice p. Moreover, the former item of Eq. (2) represents the interference from the same group, and the latter item is the interference from the users occupying the same subchannel in other groups.

Then, the data rate of user k in group g on slice C decoding the common mMTC flow can be derived as

$$r_{k,g}^c(t) = \sum_{m\in\mathcal{M}} x_g^{b,m}(t)\frac{B}{M}\log_2\left[1 + \gamma_{k,g}^{c,m}(t)\right]. \tag{3}$$

For the common mMTC data, the main objective is coverage probability. We adopt the SINR-based coverage, which means that if the user's SINR is larger than ξ, we say it is covered by the BS. Then, the coverage probability for users on slice

C at time slot t is calculated as

$$P_c^{cov}(t) = \mathbb{P}\{\gamma_{k,g}^c(t) > \xi\} = \mathbb{P}\left\{\sum_{m\in\mathcal{M}} x_g^{b,m}(t)\gamma_{k,g}^{c,m}(t) > \xi\right\}. \quad (4)$$

After decoding the common flow, it will be removed from the received signal using SIC. To simplify the analysis, it is assumed that the SIC procedure is perfect, i.e., no error propagation occurs in this paper. Then, each user can decode its own private data. Here, two-user grouping is considered, thus the received SINR of user k on slice $p \in \{u, e\}$ decoding its private flow in group g with subchannel m at time slot t can be expressed as:

$$\gamma_{k,g}^{p,m}(t) = \frac{P_{k,g}^{p,m}(t)|h_k^m(t)|^2}{I_{k,g}^{p,m}(t) + N_0}, \quad (5)$$

where $I_{k,g}^{p,m}(t)$ represents the interference of user k on slice C in group g with subchannel m, which is calculated by

$$I_{k,g}^{p,m}(t) = P_{k,g}^{\{u,e\}\setminus p,m}(t)|h_k^m(t)|^2 + \sum_{g'\in\mathcal{G}\setminus g}\sum_{i=1}^{K} x_{g'}^{b,m}(t)P_{i,g'}^m(t)|h_k^m(t)|^2, \quad (6)$$

where the former item of Eq. (6) represents the interference from the privacy data in the same group, and the latter item is the interference from the users occupying the same subchannel in other groups.

For the private URLLC data, the primary target is to reduce the service delay and improve reliability. We assume that the process of random data arrivals of private URLLC data u of user k at slot t is denoted as $A_{k_u}(t)$, which is assumed as independent and identically distributed (i.i.d.) over the slots and followed a Poisson arrival process with an arrival rate of λ_U. Firstly, we can calculate the downlink data rate of user k on slice U decoding the private URLLC[2] flow at time slot t as

$$R_k^u(t) = \sum_{g=1}^{G}\sum_{m}^{M} c_{k,g}(t)x_g^{b,m}(t)\frac{B}{M}\log_2\left[1+\gamma_{k,g}^{u,m}(t)\right]. \quad (7)$$

Then, the total average delay $D_U(t)$ of all mULC users on slice U decoding private URLLC flows can be calculated as

$$D_U(t) = \frac{1}{K_U}\sum_{k}^{K_U}\frac{A_{k_u}(t)}{R_k^u(t)}. \quad (8)$$

Additionally, the communication reliability is defined as the success connection probability [42], which means the probability of the achievable rate exceeds a pre-defined threshold. Then, the communication reliability $P_r(t)$ of slice U decoding URLLC flows can be expressed as:

$$P_r(t) = \mathbb{P}[R_k^u(t) \geq \lambda_U] = E_k\left\{P\Big\{\sum_{g=1}^{G}\sum_{m}^{M} c_{k,g}(t)x_g^{b,m}(t)\frac{B}{M}\log_2\left[1+\gamma_{k,g}^{u,m}(t)\right] \geq \lambda_U\Big\}\right\}. \quad (9)$$

[2]To simplify the analysis, we do not consider finite blocklength. Therefore, we use the asymptotic rate of URLLC.

nd delay, we defined the variable $D_U^{re}(t)$ as the reliable delay to combine the two metrics, which can be formulated as,

$$D_U^{re}(t) = \frac{D_U(t)}{P_r(t)}. \quad (10)$$

Finally, the throughput is regarded as the optimization objective for uMBB users on slice E decoding the private eMBB flows. Then, the total throughput can be expressed as

$$R_E(t) = \sum_{g=1}^{G}\sum_{k=1}^{K_M}\sum_{m}^{M} c_{k,g}(t)x_g^{b,m}(t)\frac{B}{M}\log_2\left[1+\gamma_{k,g}^{e,m}(t)\right]. \quad (11)$$

IV. Problem Formulation and Solution

In this section, the utility maximization problem of the system will be defined, which is a MINLP. To deal with MINLP, traditional optimization algorithms require a procedure of convex hull relaxations or linear approximation, such as [43], [44]. Fundamentally, these methods obtain an approximation of the MINLP rather than solve the original problem, which may not a feasible solution. DRL-based algorithms are proposed for directly solving MINLPs. In which, the independent DDPG algorithms can achieve higher convergence due to the decentralized learning. Thus, we propose a 2IADDPG algorithm to solve the above problem.

A. Utility Maximization Problem Formulation

As each hybrid service has one more QoS requirement, we define a weighted-sum utility function as:

$$U(t) = \beta_1(\psi - D_U^{re}(t)) + \beta_2 P_M^{cov}(t) + \beta_3 R_E(t), \quad (12)$$

where β_1, β_2, and β_3 are the weight of the reliable delay, coverage probability, and throughput, respectively. In addition, ψ is the initial maximum benefit of reliable delay to ensure the non-negativity of the utility [45], [46].

Then, we aim to maximize the utility function for the RSMA-enabled RAN system with the constrained resource, which can be formulated as follows:

$$\begin{aligned}
\mathbf{P1}: \quad & \max_{\mathbf{c},\mathbf{x},\mathbf{P}} \ \beta_1(\psi - D_U^{re}) + \beta_2 P_M^{cov} + \beta_3 R_E \\
\text{s.t. } & C1: x_g^{b,m}(t), c_{k,g}(t) \in \{0,1\}, \forall k, g, m, b, t, \\
& C2: \sum_{m\in\mathcal{M}} x_g^{b,m}(t) \leq 1, \forall g, b, t, \\
& C3: \sum_{g\in\mathcal{G}} c_{k,g}(t) \leq 1, \forall k, t, \\
& C4: \sum_{m\in\mathcal{M}}\left[\sum_{p\in\{e,u\}} P_{k,g}^{p,m}(t) + P_g^{c,m}(t)\right] \leq P_U, \forall k, g, t, \\
& C5: h_k^m(t)\big[P_g^{c,m}(t) - \sum_{p\in\{u,e\}} P_{k,g}^{p,m}(t)\big] - N_0 \leq \theta, \forall k, g, m, t,
\end{aligned} \quad (13)$$

where $\mathbf{P} = \Big\{P_g^{c,m}(t), P_{k,g}^{p,m}(t) | \forall k, g, c, m, t\Big\}$. $C1$ represents the value range of binary variables $x_g^{b,m}(t)$ and $c_{k,g}(t)$, while $C2$ means that we could assign at most one subchannel to a group in the same slot, and $C3$ represents the grouped users can not be selected again. $C4$ is the maximum power allocated

to the users in a group, while $C5$ is to promise successful implementation of the SIC procedure at the receiver [47] and θ is a non-negative value in dBm [48].

B. MDP Modelling

Considering the dynamic characteristics in the proposed transmission scenario, the optimization problem **P1** can be described as a Markov decision process (MDP). A MDP can be denoted by a four-tuple of $\left\langle \mathbf{O}, \mathbf{A}, \mathbf{R}, \mathbf{O}' \right\rangle$, where $\mathbf{O}$ and $\mathbf{O}'$ are the observed state at the current time slot and next time slot, $\mathbf{A}$ is the set of actions, and $\mathbf{R}$ represents the reward of the agent which can be customized according to the different system. The four-tuples will be stored in a buffer, from which a mini-batch of samples is randomly selected to train the neural network. In our model, two agents are adopted to take action. Let $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$ denote the two agents, where $\boldsymbol{A_1}$ performs user grouping and subchannel allocation, and $\boldsymbol{A_2}$ optimizes power control for users based on the selected actions a_t^1. Then, we can get the four-tuples for the first agent as below.

The state and next state of $\boldsymbol{A_1}$: In the RAN system, the states mainly include the channel states in the physical network. Then, the state $o_t^1 \in \mathbf{O_1}$ at time slot t of $\boldsymbol{A_1}$ is

$$o_t^1 = \{h_1^1(t), \cdots, h_k^m(t), \cdots, h_K^M(t)\}. \tag{14}$$

Obviously, the next state $o_{t+1}^1 \in \mathbf{O_1'}$ at time slot $t+1$ can be denoted as:

$$o_{t+1}^1 = \{h_1^1(t+1), \cdots, h_k^m(t+1), \cdots, h_K^M(t+1)\}. \tag{15}$$

Action of $\boldsymbol{A_1}$: Based on the current state of the system and the observed environment, $\boldsymbol{A_1}$ selects actions from the action space $\mathbf{A_1}$. The most suitable action space should contain all possible user grouping results and subchannel assignments, hence the set of actions at the time slot t is:

$$a_t^1 = \left\{c_{k,g}(t), x_g^{b,m}(t), \forall k, g, m, b\right\}. \tag{16}$$

As $c_{k,g}(t)$ and $x_g^{b,m}(t)$ are discrete actions, which are relaxed to continuous variables with a value range of [0,1], which conforms to the continuous action space of the proposed DDPG-based algorithm.

Reward of $\boldsymbol{A_1}$: The agent will measure the performance by a scalar reward r_t^1 to assist making better decisions for higher reward. According to (12), the reward space of $\boldsymbol{A_1}$ can be expressed as follows:

$$r_t^1 = U(t). \tag{17}$$

Similarly, we introduce the four-tuples for $\boldsymbol{A_2}$. Referring to the actions performed by the $\boldsymbol{A_1}$, the $\boldsymbol{A_2}$ optimizes power allocation only for the users assigned subchannels.

The state and next state of $\boldsymbol{A_2}$: Similarly, the state $o_t^2 \in \mathbf{O_2}$ at time slot t and the next state $o_{t+1}^2 \in \mathbf{O_2'}$ at time slot $t+1$ of $\boldsymbol{A_2}$ can be defined as:

$$\begin{cases} o_t^2 = \{c_{k,g}(t), h_1^1(t), \cdots, h_K^M(t)\}, \\ o_{t+1}^2 = \{c_{k,g}(t+1), h_1^1(t+1), \cdots, h_K^M(t+1)\}. \end{cases} \tag{18}$$

Action of $\boldsymbol{A_2}$: Correspondingly, $\boldsymbol{A_2}$ selects actions from the action space $\mathbf{A_2}$. The actions of $\boldsymbol{A_2}$ should be including all the possible power allocation decisions, then the set of actions at the time slot t is:

$$a_t^2 = \left\{P_g^{c,m}(t), P_{\bar{k},g}^{p,m}(t), \forall k, g, m, p \in \{u, e\}\right\}, \tag{19}$$

where $\bar{k}$ represents the user k who has been allocated with subchannels in step 1.

Reward of $\boldsymbol{A_2}$: The reward of $\boldsymbol{A_2}$ is defined as

$$r_t^2 = U(t). \tag{20}$$

C. Deep Deterministic Policy Gradient

In our proposed 2IADDPG framework, the two independent DDPG agents rely on the same actor-critic (AC) architecture, where each agent contains an actor network for action generation and a critic network for action evaluation, as illustrated in Fig. 2. Moreover, the policy network $\mu(o_t|\theta^\mu)$ and the action-value function $Q(o_{t+1}, \mu(o_{t+1})|\theta^Q)$ are referred to an actor network and a critic network, to train the parameters θ^μ and θ^Q, respectively. Moreover, both the critic network and the actor network are also created with a copy: the target actor network and the target critic network, which are created to slowly update the learned actor and critic network by significantly increasing training stability.

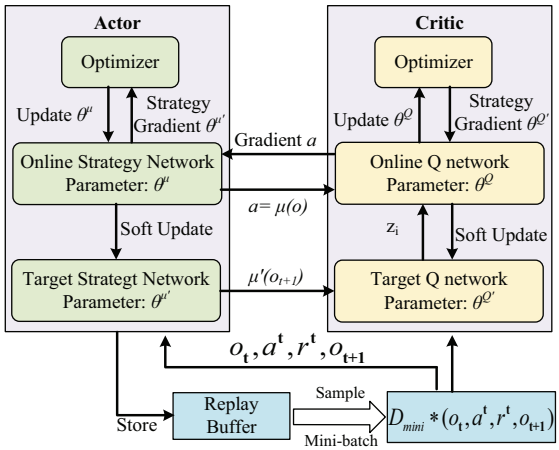

Fig. 2. The framework of the DDPG model.

Specifically, the four components are described as:

1) *Online Critic Network:* Parameterized by θ^Q to evaluate the action by a state-action value function $Q(o_t, a_t|\theta^Q)$, which can be calculated with the Bellman equation as

$$Q(o_t, a_t|\theta^Q) = \mathbb{E}[r(o_t, a_t) + \gamma Q(o_{t+1}, \mu(o_{t+1})|\theta^Q)], \tag{21}$$

where $\gamma \in (0, 1)$ is the discount factor, and $r(o_t, a_t)$ is the reward function defined in section IV-B. The parameter θ^Q is updated by minimizing the loss function, which is defined as the mean-squared Bellman error as

$$L(\theta^Q) = E[(Q(o_t, a_t|\theta^Q) - y_t)^2)], \tag{22}$$

where $y_t = r(o_t, a_t) + \gamma Q(o_{t+1}, \mu(t+1)|\theta^Q)$ is the target state-action value. And the gradient of $L(\theta^Q)$ is defined as

$$\nabla_{\theta^Q} L = \mathbb{E}\left[2(y_t - \nabla_{\theta^Q}(o_t, a_t)Q(o_t, a_t|\theta^Q))\right]. \tag{23}$$

In addition, a mini-batch with size D_{mini} will be sampled from the replay buffer for updating the parameter θ^Q by the stochastic gradient descent method, which can be given by

$$\theta^Q = \theta^Q - \frac{\alpha_Q}{D_{mini}} \sum_T \left[2(y_t - \nabla_{\theta^Q} Q(o_t, a_t|\theta^Q))\right], \quad (24)$$

where α_Q denotes the learning rate of the critic network.

2) *Online Actor Network:* Parameterized by θ^μ to take action a_{t+1} based on the state information o_{t+1} by sampling a mini-batch uniformly from the replay buffer. The action generation policy $\mu(o_t|\theta^\mu)$ is updated by using the gradient descent method as

$$\nabla_{\theta^\mu} J \approx E\left[\nabla_a Q(o_t, \mu(o_t|\theta^\mu)|\theta^Q) \cdot \nabla_{\theta^\mu} \mu(o_t|\theta^\mu)\right]. \quad (25)$$

Moreover, the parameter θ^μ is updated by

$$\theta^\mu = \theta^\mu - \frac{\alpha_\mu}{D_{mini}} \sum_T \left[\nabla_a Q(o_t, \mu(o_t|\theta^\mu)|\theta^Q) \cdot \nabla_{\theta^\mu} \mu(o_t|\theta^\mu)\right], \quad (26)$$

where α_μ denotes the learning rate of the actor network.

3) *Target Critic Network:* Updating weight $\theta^{Q'}$ of the value network and then giving the current Q value.

4) *Target Actor Network:* Calculating the current value Q', as well as obtaining the target value.

The target critic network and actor network are parameterized by $\theta^{Q'}$ and $\theta^{\mu'}$, respectively. $\theta^{Q'}$ and $\theta^{\mu'}$ are updated by the soft update method with the constant τ as follows

$$\begin{cases} \theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}, \\ \theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}, \end{cases} \quad (27)$$

where the default value τ can range from 0.0001 to 0.1.

D. The proposed 2IADDPG algorithm

In this section, we customize a 2IADDPG model and propose a 2IADDPG algorithm to solve the above problem. The global framework of the algorithm is illustrated in Fig. 3, which is composed of two independent agents to realize independent decision-making. Considering the independent relationship between the user grouping, subchannel allocation, and power allocation optimization, we can divide the total action space into two sub-spaces, which are optimized by two agents. To elaborate, the first agent (referring Agent $\boldsymbol{A_1}$) optimizes the user grouping and subchannel allocation, then the second agent (referring Agent $\boldsymbol{A_2}$) is responsible for power control based on the results of Agent $\boldsymbol{A_1}$. In addition, the selected actions will be judged to meet the constraints of C2-C5, if the constraints are not satisfied, the action will be re-selected. In addition, we have added the random process $\mathcal{N}_t^i$ for action exploration to avoid infinite loops of action re-selections. Finally, the utility maximization problem can be solved by the two-step iteration process. The specific process is summarized in Algorithm 1.

Moreover, each agent in the proposed 2IADDPG algorithm uses a fully connected deep neural network (DNN) structure. For the actor 1, we built a DNN with $K \times M$ inputs and $K \times M$ outputs, and the critic 1 is set up by a DNN with $2K \times M$ inputs and 1 output. Similarly, for the actor 2, we

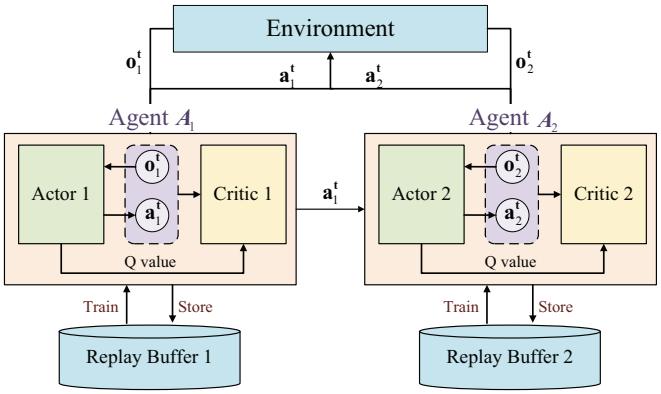

Fig. 3. The structure of the proposed 2IADDPG algorithm.

Algorithm 1 The proposed 2IADDPG algorithm

1: Randomly initialize critic network Q_i and actor network μ_i with weights $\theta_i^{Q_i}$ and $\theta_i^{\mu_i}$, $i \in \{1, 2\}$.
2: Initialize target critic network Q_i' and actor network μ_i' with weights $\theta_i^{Q_i'} \longleftarrow \theta_i^{Q_i}$ and $\theta_i^{\mu_i'} \longleftarrow \theta_i^{\mu_i}$, $i \in \{1, 2\}$.
3: Initialize replay buffer D_1 and D_2.
4: **for** $t = 0, 1, 2, \ldots, T-1$ **do**
5: Initialize a random process $\mathcal{N}_t^i$ for action exploration;
6: Obtain the observation state o_t^1.
7: For the Agent $\boldsymbol{A_1}$, choose action $a_t^1 = \mu_1(o_t^1|\theta_1^{\mu_1}) + \mathcal{N}_t^1$ with the deterministic policy;
8: Obtain observation state o_t^2;
9: For the Agent $\boldsymbol{A_2}$, select action $a_t^2 = \mu_2(o_t^2|\theta_2^{\mu_2}) + \mathcal{N}_t^2$ based on the current state and policy;
10: Evaluate whether the chosen action satisfies the constraint conditions. If not, the agents will re-choose the action;
11: Obtain the reward $r_t^i(o_t^i, a_t^i)$ and the next state o_{t+1}^i, for $i \in \{1, 2\}$;
12: **if** the replay buffer is not full **then**
13: Store $\left\langle o_t^i, a_t^i, r_t^i, o_{t+1}^i \right\rangle$ in replay buffer D_i;
14: **else**
15: Sample a random mini-batch of D_{mini} transitions $\left\langle o_j^i, a_j^i, r_j^i, o_{j+1}^i \right\rangle$ from the replay buffer D_i;
16: Calculate $y_j^i = r_j^i + \gamma_i Q_i'(o_{j+1}^i, \mu_i'(o_{j+1}^i|\theta_i^{\mu_i'})|\theta_j^{Q_i'})$, for $i \in \{1, 2\}$;
17: Update the critic network by minimizing the loss function in (22);
18: Update the actor network by sampled stochastic policy gradient ascent with (25);
19: Update the parameters of the target actor network and the target critic network according to (27);
20: **end if**
21: $t \leftarrow t + 1$.
22: **end for**

set up a DNN with $K \times M$ inputs and $\frac{3K}{2}M$ outputs, and the critic network is set up by a DNN with $\frac{5}{2}K \times M$ inputs and 1 output. Moreover, there is no known rule for determining the number of hidden layers and neurons. It is appropriate to select the parameters through a trial-and-error approach [49], [50]. The specific configuration of the DNNs is summarized in Table I.

TABLE I: DNNs configuration for the 2IADDPG algorithm

	Number of hidden layers	Number of neurous	Activiation function
Actor i	3	256 + 256 + 8	Relu + Relu + Tanh
Critic i	2	256 + 8	Relu + Relu

V. Simulation Results And Discussions

In this section, we present the extensive set of simulation results to evaluate the theoretical analysis and compare the obtained performance with the benchmark schemes. We suppose the fixed size of the experience memory D_i as 5000 four-tuples for each agent, and the mini-batch size is defined as D_{mini}=32 for training at each time step. According to [34], [49], the other parameters of the simulation are summarized in Table II.

TABLE II: Parameter Settings

Parameter	Description	Value
N	Number of BSs	2
B	Total system bandwidth	10 MHz
M	Number of subchannels	10
N_0	Noise power	-90 dBm/Hz
λ_U	Random data arrival rate of URLLC data	50 kbits/slot
ξ	SINR threshold of the coverage probability	0.15
P_B	Maximum transmission power of each BS	40W
P_U	Maximum power allocated to one group	5W
ψ	Initial maximum benefit of the reliable delay	0.2
β_1	Weight of the reliable delay	50
β_2	Weight of the coverage probability	10
β_3	Weight of the throughput	0.00000005
α_μ	Learning rate of actor network	0.0001
α_Q	Learning rate of critic network	0.0002
γ	Discount factor	0.9

In order to show the advantages of our proposed schemes, we include the five benchmark schemes for comparison.

- Benchmark 1 is the orthogonal frequency-division multiple access (OFDMA) scheme as the comparison of the proposed RSMA scheme: the users requesting different services will be allocated with orthogonal subchannels.
- Benchmark 2 is the non-orthogonal multiple access (NOMA) method. In the NOMA scheme, the mixed traffic of each user is not split and the data of users in the same NOMA pair is encoded into a single data stream. Accordingly, two slices are constructed in NOMA, i.e., mULC and uMBB slices, to accommodate the mixed traffic. The users are paired into multiple groups, and we consider the case of two-user grouping, i.e., a mULC user and a uMBB user are included in a group and are transmitted non-orthogonally. Among them, the user grouping is optimal, and after the pairing is completed, the power allocation is determined according to the user's channel conditions. Specifically, the user with good channel condition (strong user) will be allocated lower power, and the user with weak channel condition (weak user) will be assigned higher power.
- Then, in Benchmark 3, we compare the proposed 2IADDPG algorithm with the "MADDPG [51]" algorithm, and the number of the agents is set to 2.
- Next, the conventional single-agent "DDPG [34]" algorithm is Benchmark 4, as the comparison of the proposed 2IADDPG algorithm.
- Further, we consider the "A3C [32]" algorithm to be Benchmark 5 of the proposed 2IADDPG algorithm.
- Finally, Benchmark 6 is the "PPO [31]" algorithm.

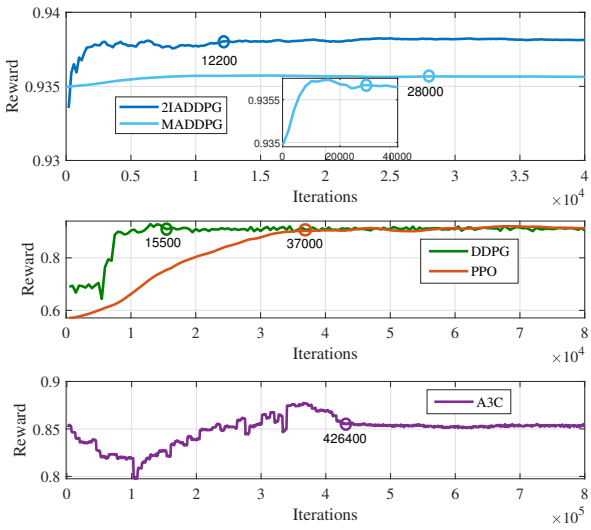

Fig. 4. Convergence of the 2IADDPG algorithm versus the benchmarks.

Fig. 4 and Fig. 5 show the convergence performance of the proposed algorithm versus the benchmarks and different numbers of users, respectively. As shown in Fig. 4, our proposed algorithm shows a faster convergence than the shown competitors, the reward converges to a relatively stable state after about 12200 training iterations. While the MADDPG algorithm and DDPG algorithm converge after about 28000 and 15500 training iterations. Moreover, the PPO algorithm and A3C algorithm converge after about 37000 and 426400 iterations, respectively. In the MADDPG algorithm, due to the sharing memory and the possible policy interaction, there will be a problem of policy interference, that is, the update of one agent may negatively affect the performance of other agents. In the proposed 2IADDPG algorithm, the agents can explore their own strategies more efficiently because each agent

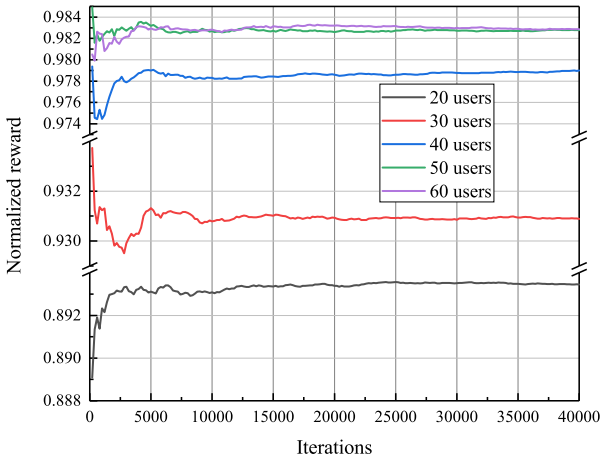

Fig. 5. Convergence of the 2IADDPG algorithm with different number of users.

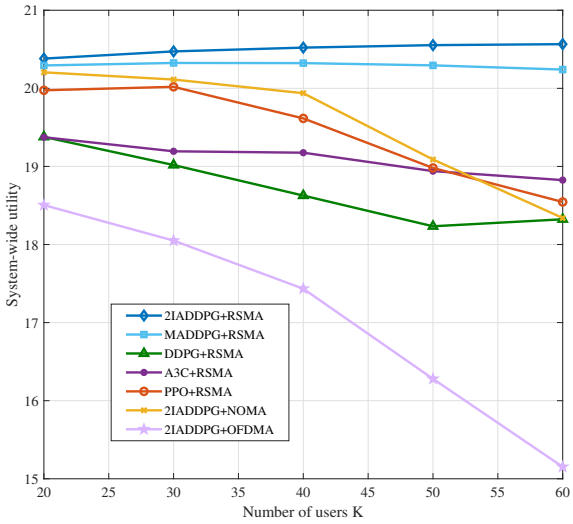

Fig. 6. System-wide utility versus the number of users.

independently generates actions and updates its strategies. Thus, our proposed 2IADDPG algorithm can explore better strategies faster and achieve faster convergence. Additionally, the number of users K of the system also affects the reward values as in Fig. 5. As the number of users increases, the system-wide utility increases, which corresponds to the growth of reward.

Fig. 6 characterizes several utility curves versus the number of users for the proposed scheme and other comparison schemes. We can see that the utility of the "2IADDPG+RSMA" scheme is increasing when the number of users increases, while the utilities of the other comparison schemes are decreasing when the number of users increases. As the number of users increases, the available subchannels of the OFDMA scheme will decrease, then there will be more users who may not be allocated with subchannels. As a result, the performance of users will be degraded, and thus the utility will be reduced. In addition, the RSMA and NOMA schemes are superior to the OFDMA scheme due to the SIC receivers. More importantly, the RSMA scheme surpasses NOMA and OFDMA schemes by adjusting the splitting power of two flows for each user and enabling partial decoding of interference as well as treating part of the remaining interference as noise, so as to control the decoding interference. As expected, our proposed 2IADDPG algorithm can achieve higher utility than the MADDPG, DDPG, and other comparison algorithms. It indicates that MNO can achieve higher utility at higher traffic loads by our proposed 2IADDPG algorithm and RSMA scheme, which means the users can get higher performance with our schemes.

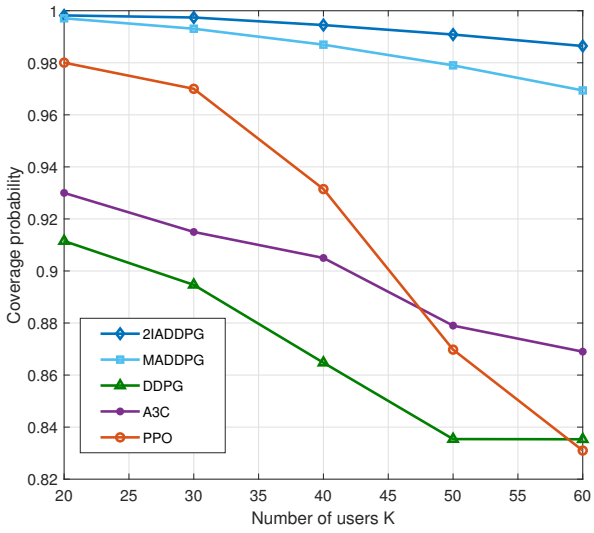

Fig. 7. Coverage probability versus the number of users K on mMTC slice.

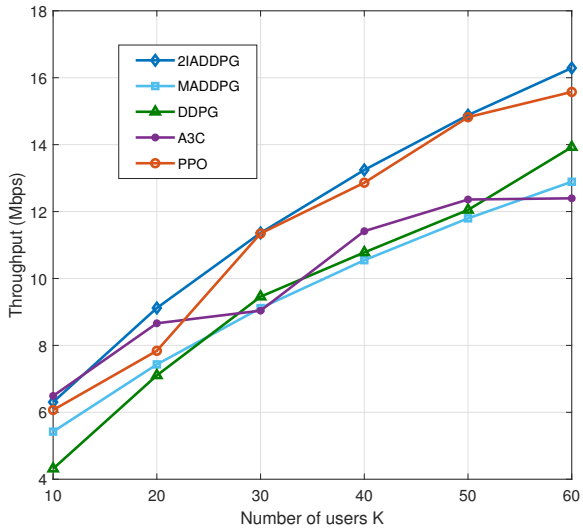

Fig. 8. Throughput versus the number of users K on eMBB slice.

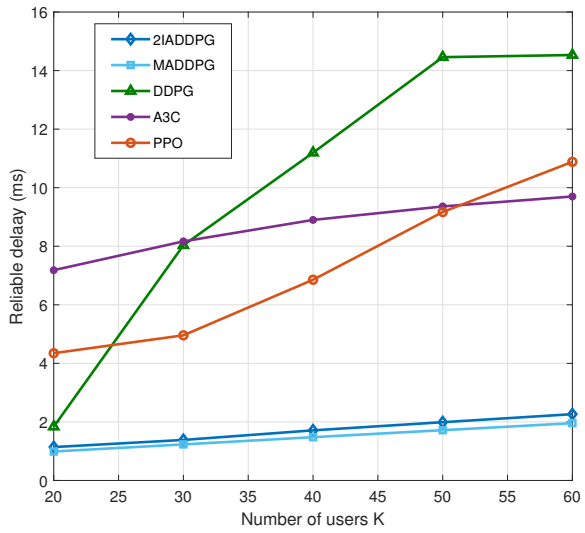

Fig. 9. Reliable delay versus the number of users K on URLLC slice.

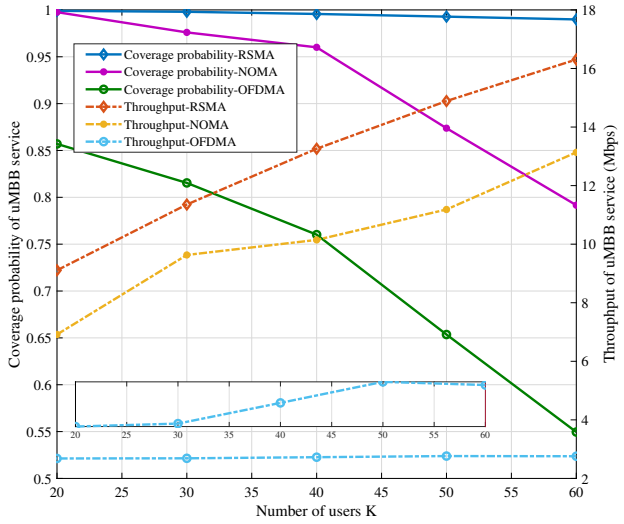

Fig. 10. Performance versus the number of users of uMBB service.

Fig. 7 exhibits the coverage probability versus the number of users K on slice C for the proposed and other comparison algorithms. We can see that the coverage probability of all schemes decreases with the increase of the number of users. This is because as the number of users increases, resource competition becomes fierce, and interference increases, which will lead to degradation in coverage performance. Compared with the MADDPG, DDPG, PPO, and A3C algorithms, our proposed 2IADDPG algorithm can achieve higher coverage.

Then, the trends of throughput and reliable delay of the corresponding slices for the proposed and other comparison algorithms are shown in Fig. 8 and Fig. 9, respectively. Similarly, the 2IADDPG algorithm always shows better performances than the mentioned benchmark algorithms. Though, the MADDPG achieves a slightly lower reliable delay than the proposed algorithm as shown in Fig. 9, the throughput performance is much lower than the proposed 2IADDPG algorithm in Fig. 8. Moreover, the throughput of all the schemes in Fig. 8 increases with the increase of the number of users. We also note that the reliable delay in Fig. 9 increases with the increase of the number of users, this is because the increase of the number of users leads to greater resource competition and lower resource allocation to users, which results in the increase of reliable delay.

Fig. 10 depicts the coverage probability and throughput of uMBB service versus the number of users K. Meanwhile, Fig. 11 portrays the coverage probability and reliable delay of mULC service versus the number of users. Upon increasing K, the coverage probabilities of uMBB and mULC services decrease, while the throughput of the uMBB service and the reliable delay of mMTC service increases. Compared with the orthogonal scheme, the non-orthogonal schemes can significantly improve the coverage probability, as the non-orthogonal schemes allow subchannels sharing within multiple users, thus providing a higher access possibility. Importantly,

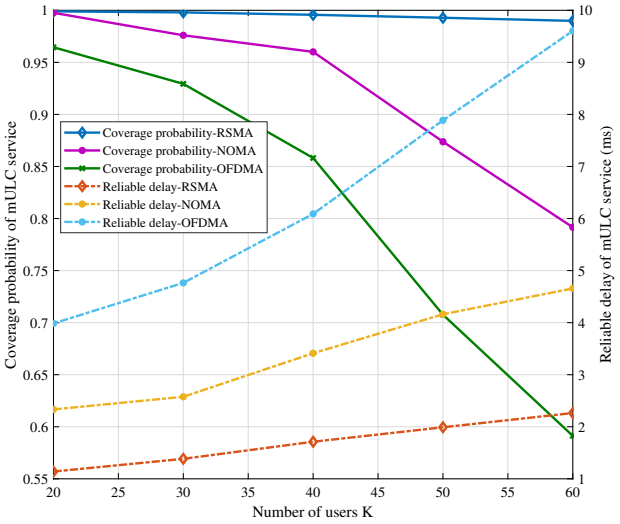

Fig. 11. Performance versus the number of users of mULC service.

the RSMA scheme can always achieve higher performances than the NOMA and OFDMA schemes.

VI. Conclusion

In this paper, we propose an original RSMA-enabled hybrid RAN slicing scheme. Firstly, the basic URLLC, eMBB, and mMTC slices are constructed to cooperatively provide hybrid services. Then, the uMBB and mULC users are assumed to share the same subchannels by using RSMA, in which the data of each user will be divided into the common mMTC data and the private eMBB/URLLC data. In this way, the users can connect with two basic slices to obtain services. In addition, a two-user grouping solution is adopted in this

paper, i.e., users are divided into several disjoint RSMA groups and each group includes one uMBB user and one mULC user. Furthermore, the utility maximization problem is formulated to jointly optimize the heterogeneous performance metrics, which is defined as the sum-weight of reliable delay, throughput, and coverage probability. Finally, an original 2IADDPG framework is customized, and a 2IDDPG algorithm is proposed to solve the target problem, wherein the first agent is employed to obtain user grouping and subchannel allocation, while the second agent is applied for power control based on the results of the first agent.

The encouraging results achieved so far are pushing us to plan for additional related research work in the future. In particular, we are working on the integration with other key deployment scenarios. Given the characteristics of RSMA, our future work will focus on balancing security and reliability. We aim to harness the decoding reliability offered by priority encoding strategies while addressing the potential security risks associated with shared codebook theft. By comprehensively optimizing the transmission scheme, we will ensure both the security and reliability of data transmission, thereby providing a more robust solution for future communication systems. As fairness is another critical factor in resource allocation, our future work will focus on integrating fairness considerations and exploring advanced resource allocation strategies to ensure a more balanced distribution among users while maintaining overall network efficiency. Specifically, we will investigate fairness-aware utility function design, such as integrating proportional fairness and introducing adaptive weight adjustment mechanisms, to balance the resource allocation and cope with diverse QoS demands.

REFERENCES

[1] H. Xiang, S. Yan, and M. Peng, "A Realization of Fog-RAN Slicing via Deep Reinforcement Learning," *IEEE Transactions on Wireless Communications*, vol. 19, no. 4, pp. 2515–2527, 2020.

[2] M. Rasti, S. K. Taskou, H. Tabassum, and E. Hossain, "Evolution Toward 6G Multi-Band Wireless Networks: A Resource Management Perspective," *IEEE Wireless Communications*, vol. 29, no. 4, pp. 118–125, 2022.

[3] C.-X. Wang, X. You, X. Gao, X. Zhu, Z. Li, C. Zhang, H. Wang, Y. Huang, Y. Chen, H. Haas, J. S. Thompson, E. G. Larsson, M. D. Renzo, W. Tong, P. Zhu, X. Shen, H. V. Poor, and L. Hanzo, "On the Road to 6G: Visions, Requirements, Key Technologies, and Testbeds," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 905–974, 2023.

[4] S. R. Pokhrel, J. Ding, J. Park, O.-S. Park, and J. Choi, "Towards Enabling Critical mMTC: A Review of URLLC Within mMTC," *IEEE Access*, vol. 8, pp. 131 796–131 813, 2020.

[5] H. Ding, X. Li, Y. Cai, B. Lorenzo, and Y. Fang, "Intelligent Data Transportation in Smart Cities: A Spectrum-Aware Approach," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2598–2611, 2018.

[6] X. Zhang, Q. Zhu, and H. V. Poor, "Neyman-Pearson Criterion Driven NFV-SDN Architectures and Optimal Resource-Allocations for Statistical-QoS Based mURLLC Over Next- Generation Metaverse Mobile Networks Using FBC," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 3, pp. 570–587, 2024.

[7] M. Zambianco and G. Verticale, "Interference Minimization in 5G Physical-Layer Network Slicing," *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4554–4564, 2020.

[8] Y. Mao, O. Dizdar, B. Clerckx, R. Schober, P. Popovski, and H. V. Poor, "Rate-Splitting Multiple Access: Fundamentals, Survey, and Future Research Trends," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 4, pp. 2073–2126, 2022.

[9] S. Zhang, B. Clerckx, D. Vargas, O. Haffenden, and A. Murphy, "Rate-Splitting Multiple Access: Finite Constellations, Receiver Design, and SIC-free Implementation," *IEEE Transactions on Communications*, pp. 1–1, 2024.

[10] R. Huang, V. W. Wong, and R. Schober, "Rate-Splitting for Intelligent Reflecting Surface-Aided Multiuser VR Streaming," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 5, pp. 1516–1535, 2023.

[11] Y. Mao, B. Clerckx, and V. O. Li, "Rate-splitting Multiple Access for Downlink Communication Systems: Bridging, Generalizing, and Outperforming SDMA and NOMA," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, p. 133, 2018.

[12] G. Zhou, L. Zhao, K. Liang, G. Zheng, and L. Hanzo, "Utility Analysis of Radio Access Network Slicing," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 1163–1167, 2020.

[13] H. Gu, L. Zhao, Z. Han, G. Zheng, and S. Song, "AI-Enhanced Cloud-Edge-Terminal Collaborative Network: Survey, Applications, and Future Directions," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 2, pp. 1322–1385, 2024.

[14] T. T. Nguyen, N. D. Nguyen, and S. Nahavandi, "Deep Reinforcement Learning for Multiagent Systems: A Review of Challenges, Solutions, and Applications," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3826–3839, 2020.

[15] A. Tampuu, T. Matiisen, D. Kodelja, I. Kuzovkin, K. Korjus, J. Aru, J. Aru, and R. Vicente, "Multiagent Cooperation and Competition with Deep Reinforcement Learning," *PloS one*, vol. 12, no. 4, p. e0172395, 2017.

[16] G. Zhou, L. Zhao, G. Zheng, S. Song, J. Zhang, and L. Hanzo, "Multiobjective Optimization of Space–Air–Ground-Integrated Network Slicing Relying on a Pair of Central and Distributed Learning Algorithms," *IEEE Internet of Things Journal*, vol. 11, no. 5, pp. 8327–8344, 2024.

[17] M. Setayesh, S. Bahrami, and V. W. Wong, "Resource Slicing for eMBB and URLLC Services in Radio Access Network Using Hierarchical Deep Learning," *IEEE Transactions on Wireless Communications*, vol. 21, no. 11, pp. 8950–8966, 2022.

[18] J. J. Alcaraz, F. Losilla, A. Zanella, and M. Zorzi, "Model-Based Reinforcement Learning with Kernels for Resource Allocation in RAN Slices," *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 486–501, 2023.

[19] K. Yu, H. Zhou, Z. Tang, X. Shen, and F. Hou, "Deep Reinforcement Learning-Based RAN Slicing for UL/DL Decoupled Cellular V2X," *IEEE Transactions on Wireless Communications*, vol. 21, no. 5, pp. 3523–3535, 2022.

[20] N. A. Khan and S. Schmid, "AI-RAN in 6G Networks: State-of-the-Art and Challenges," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 294–311, 2024.

[21] J. Zeng, T. Wu, Y. Song, Y. Zhong, T. Lv, and S. Zhou, "Achieving Energy-Efficient Massive URLLC Over Cell-Free Massive MIMO," *IEEE Internet of Things Journal*, vol. 11, no. 2, pp. 2198–2210, 2024.

[22] X. Zhang, J. Wang, and H. V. Poor, "Statistical Delay and Error-Rate Bounded QoS Provisioning for mURLLC Over 6G CF M-MIMO Mobile Networks in the Finite Blocklength Regime," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 3, pp. 652–667, 2021.

[23] S. K. Singh, K. Agrawal, K. Singh, and C.-P. Li, "Ergodic Capacity and Placement Optimization for RSMA-Enabled UAV-Assisted Communication," *IEEE Systems Journal*, vol. 17, no. 2, pp. 2586–2589, 2023.

[24] H. Xia, X. Zhou, S. Han, and C. Li, "Security-Reliability Tradeoff in RSMA-Based Communications Against Eavesdropper Collusion," *IEEE Wireless Communications Letters*, vol. 12, no. 9, pp. 1504–1507, 2023.

[25] Y. Mao, E. Piovano, and B. Clerckx, "Rate-Splitting Multiple Access for Overloaded Cellular Internet of Things," *IEEE Transactions on Communications*, vol. 69, no. 7, pp. 4504–4519, 2021.

[26] H. Cho, B. Ko, B. Clerckx, and J. Choi, "Coverage Increase at THz Frequencies: A Cooperative Rate-Splitting Approach," *IEEE Transactions on Wireless Communications*, vol. 22, no. 12, pp. 9821–9834, 2023.

[27] J. Xu, O. Dizdar, and B. Clerckx, "Rate-Splitting Multiple Access for Short-Packet Uplink Communications: A Finite Blocklength Analysis," *IEEE Communications Letters*, vol. 27, no. 2, pp. 517–521, 2023.

[28] E. J. D. Santos, R. D. Souza, and J. L. Rebelatto, "Rate-Splitting Multiple Access for URLLC Uplink in Physical Layer Network Slicing With eMBB," *IEEE Access*, vol. 9, pp. 163 178–163 187, 2021.

[29] Y. Liu, B. Clerckx, and P. Popovski, "Network Slicing for eMBB, URLLC, and mMTC: An Uplink Rate-Splitting Multiple Access Approach," *IEEE Transactions on Wireless Communications*, vol. 23, no. 3, pp. 2140–2152, 2024.

[30] W. Chen, X. Qiu, T. Cai, H.-N. Dai, Z. Zheng, and Y. Zhang, "Deep Reinforcement Learning for Internet of Things: A Comprehensive Sur-

vey," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1659–1692, 2021.

[31] S. Jiang, J. Zheng, F. Yan, and S. Zhao, "Reinforcement-Learning-Based Network Slicing and Resource Allocation for Multi-Access Edge Computing Networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 10, no. 3, pp. 1132–1145, 2024.

[32] Y. Azimi, S. Yousefi, H. Kalbkhani, and T. Kunz, "Energy-Efficient Deep Reinforcement Learning Assisted Resource Allocation for 5G-RAN Slicing," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 1, pp. 856–871, 2022.

[33] J. Mei, X. Wang, K. Zheng, G. Boudreau, A. B. Sediq, and H. Abou-Zeid, "Intelligent Radio Access Network Slicing for Service Provisioning in 6G: A Hierarchical Deep Reinforcement Learning Approach," *IEEE Transactions on Communications*, vol. 69, no. 9, pp. 6063–6078, 2021.

[34] Y. Wang, L. Zhao, X. Chu, S. Song, Y. Deng, A. Nallanathan, and K. Liang, "Deep Reinforcement Learning-Based Optimization for End-to-End Network Slicing With Control- and User-Plane Separation," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 11, pp. 12 179–12 194, 2022.

[35] M. Tan, "Multi Agent Reinforcement Learning Independent vs Cooperative Agents," 2003. [Online]. Available: https://api.semanticscholar.org/CorpusID:260435822

[36] S. Hwang, H. Kim, H. Lee, and I. Lee, "Multi-Agent Deep Reinforcement Learning for Distributed Resource Management in Wirelessly Powered Communication Networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 14 055–14 060, 2020.

[37] G. O. Boateng, G. Sun, D. A. Mensah, D. M. Doe, R. Ou, and G. Liu, "Consortium Blockchain-Based Spectrum Trading for Network Slicing in 5G RAN: A Multi-Agent Deep Reinforcement Learning Approach," *IEEE Transactions on Mobile Computing*, vol. 22, no. 10, pp. 5801–5815, 2023.

[38] A. Andreou and C. X. Mavromoustakis, "6G$^+$ Networks Through Enhanced Efficiency and Sustainability With MADDPG-Driven Network Slicing in SoS Environments," *IEEE Transactions on Green Communications and Networking*, vol. 8, no. 4, pp. 1752–1761, 2024.

[39] K. M. Lee, S. G. Subramanian, and M. Crowley, "Investigation of Independent Reinforcement Learning Algorithms in Multi-agent Environments," *Frontiers in Artificial Intelligence*, vol. 5, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:240354590

[40] H. Hu, D. Shi, H. Yang, Y. Peng, Y. Zhou, and S. Yang, "Independent Multi-agent Reinforcement Learning Using Common Knowledge," in *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2022, pp. 2703–2708.

[41] P. Doanis, T. Giannakas, and T. Spyropoulos, "Scalable end-to-end slice embedding and reconfiguration based on independent DQN agents," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, 2022, pp. 3429–3434.

[42] Y. Zhang, L. Zhao, G. Zheng, X. Chu, Z. Ding, and K.-C. Chen, "Resource Allocation for Open-Loop Ultra-Reliable and Low-Latency Uplink Communications in Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 3, pp. 2590–2604, 2021.

[43] A. R. Hossain and N. Ansari, "Priority-Based Downlink Wireless Resource Provisioning for Radio Access Network Slicing," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 9273–9281, 2021.

[44] J. Shi, X. Chen, N. Huang, H. Jiang, Z. Yang, and M. Chen, "Power-Efficient Transmission for User-Centric Networks With Limited Fronthaul Capacity and Computation Resource," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5649–5660, 2020.

[45] M. Xiao, J. Wu, C. Liu, and L. Huang, "TOUR: Time-sensitive Opportunistic Utility-based Routing in delay tolerant networks," in *2013 Proceedings IEEE INFOCOM*, 2013, pp. 2085–2091.

[46] Y. Wang, L. Zhao, X. Chu, S. Song, Y. Deng, A. Nallanathan, and G. Zhou, "Two-timescale Optimization for E2E Network Slicing-aided Cloud-edge Collaborative Networks," *IEEE Transactions on Vehicular Technology*, pp. 1–13, 2025.

[47] Z. Yang, M. Chen, W. Saad, and M. Shikh-Bahaei, "Optimization of Rate Allocation and Power Control for Rate Splitting Multiple Access (RSMA)," *IEEE Transactions on Communications*, vol. 69, no. 9, pp. 5988–6002, 2021.

[48] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-Splitting for Multi-Antenna Non-Orthogonal Unicast and Multicast Transmission: Spectral and Energy Efficiency Analysis," *IEEE Transactions on Communications*, vol. 67, no. 12, pp. 8754–8770, 2019.

[49] G. Zhou, L. Zhao, G. Zheng, Z. Xie, S. Song, and K.-C. Chen, "Joint Multi-Objective Optimization for Radio Access Network Slicing Using Multi-Agent Deep Reinforcement Learning," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 9, pp. 11 828–11 843, 2023.

[50] G. Sun, Z. T. Gebrekidan, G. O. Boateng, D. Ayepah-Mensah, and W. Jiang, "Dynamic Reservation and Deep Reinforcement Learning Based Autonomous Resource Slicing for Virtualized Radio Access Networks," *IEEE Access*, vol. 7, pp. 45 758–45 772, 2019.

[51] Y. Zhang, Z. Mou, F. Gao, J. Jiang, R. Ding, and Z. Han, "UAV-Enabled Secure Communications by Multi-Agent Deep Reinforcement Learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 11 599–11 611, 2020.