Coexistence of eMBB and URLLC in Open Radio Access Networks: A Distributed Learning Framework

Madyan Alsenwi, Eva Lagunas, Symeon Chatzinotas

Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg
Luxembourg, L-1855, Luxembourg
Email: madyan.alsenwi@uni.lu, eva.lagunas@uni.lu, symeon.chatzinotas@uni.lu

***Abstract*—This paper proposes a distributed learning framework for network slicing in multi-cell open radio access networks providing two services: Ultra-Reliable Low Latency Communications (URLLC) and enhanced Mobile BroadBand (eMBB). In particular, a resource allocation optimization problem is formulated with an objective to maximize the average eMBB data rate while considering URLLC constraints and the data rate variance among eMBB users. A multi-agent Deep Reinforcement Learning (DRL) based algorithm is developed to solve the formulated problem, where network components collaboratively train a global machine learning model and then share learning parameters for distributed executions at network edges. Specifically, DRL agents are installed at Near-Real-Time Radio access network Intelligent Controllers (Near-RT RICs) located in the network edge servers to provide online resource allocation decisions while the training process is performed offline at the Non-Real-Time RIC (Non-RT RIC) located in a regional cloud server. The achieved simulation results show that the proposed algorithm can ensure the required URLLC reliability while keeping the Quality-of-Service (QoS) requirements of the eMBB service.**

***Index Terms*—O-RAN, distributed learning, DRL, network slicing, eMBB, URLLC, 5G NR.**

I. Introduction

The Open Radio Access Networks (O-RAN) Alliance, a consortium of industry and academic institutions, has introduced a new vision for Next-Generation (NextG) cellular systems, where standardized interfaces are proposed to allow operators to use shared infrastructure belonging to multiple vendors. An important innovation proposed by the O-RAN Alliance is the RAN Intelligent Controller (RIC), which enables RAN optimization via closed-control loops. Specifically, two types of RIC are introduced in the O-RAN vision [1]: 1) Non-Real-Time RIC (Non-RT RIC) and 2) Near-Real-Time RIC (Near-RT RIC). The Near-RT RIC handles operations at small time scales and enables intelligence in the RAN by hosting third-party applications (xApps), while the Non-RT RIC conducts functions with large time scales, i.e., training machine learning models [2].

Furthermore, NextG cellular networks enable diverse applications that come under three categories: 1) Ultra-Reliable Low-Latency Communications (URLLC), 2) enhanced Mobile Broad-Band (eMBB), and 3) massive Machine-Type Communications (mMTC). Practically, applications supported by the URLLC service transmit short packets sporadically. Contrarily, eMBB transmissions spread over long time intervals to improve spectral efficiency. The 3GPP has proposed the preemption (puncturing) multiplexing technique[1] that allows scheduling URLLC packets over eMBB transmissions to satisfy URLLC latency while improving spectral efficiency. Finally, mMTC devices transmit at a fixed rate; hence, statically allocating specific radio channels for MMTC service is more efficient than dynamic allocation [4].

The coexistence of eMBB and URLLC on the same radio resources with different Quality of Service (QoS) requirements leads to a challenging resource allocation problem. Recently, research on the coexistence problem of eMBB and URLLC traffics has gained attention. The work in [5] studies the impact of the puncturing resource allocation approach on eMBB service. The authors modeled this impact as convex, threshold, and linear functions. The coexistence of visual and haptic applications over wireless systems was discussed in [6]. In this study, the visual transmissions are accommodated by the eMBB service, while the haptic is considered a URLLC application. In [7], the resource allocation to eMBB and URLLC traffics is performed over two different time scales, i.e., time-slots and mini-slots. Specifically, an algorithm based on the BSUM technique was developed to allocate resources to eMBB users over time slots, while the transportation model was used to solve the URLLC scheduling problem on a mini-slots time scale. In [8], a relaxation-based algorithm was developed to solve the eMBB/URLLC resource allocation problem aiming at maximizing the network throughput. The study in [9] modeled the resource allocation problem of eMBB and URLLC services in a single cell network scenario as a risk-aware optimization problem that considers the randomization nature of URLLC traffic. The authors proposed an optimization-aided DRL algorithm to solve the formulated problem, where the training and execution stages are performed centrally at the BS.

Literature review shows that there is still a lack of studies on the eMBB-URLLC coexistence problem in multi-cell O-

[1] 3GPP has proposed a new scheduling mechanism named puncturing, also called preemptive, for dynamic multiplexing of eMBB and URLLC traffics [3]. In this approach, URLLC traffic is scheduled over short transmission time intervals on top of the ongoing eMBB transmissions by allocating zero transmission power to the selected eMBB users.

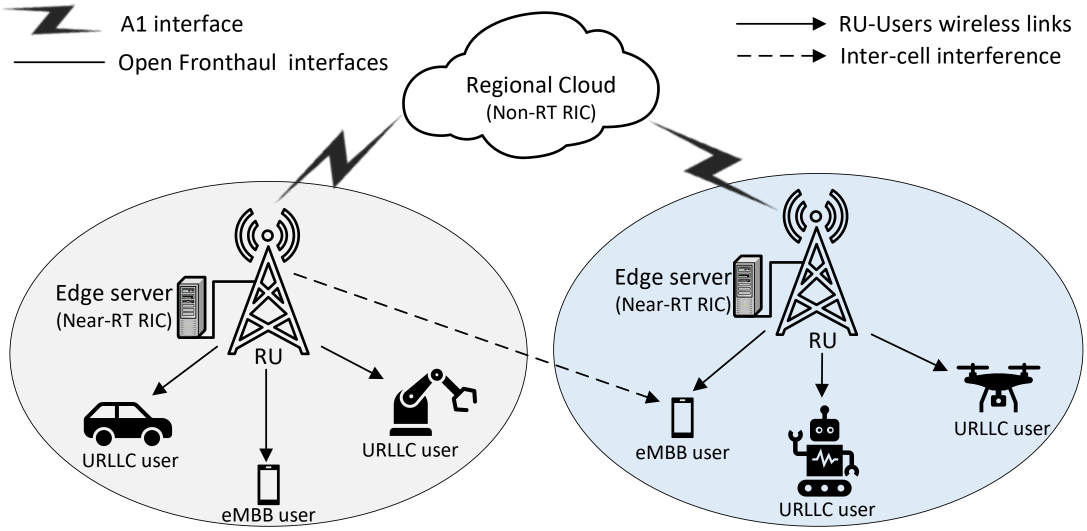

Figure 1: System model.

RANs. In this work, we leverage the advantages of O-RAN architectures in implementing learning algorithms to design a novel algorithm based on the multi-agent DRL technique to solve the dynamic multiplexing problem of eMBB and URLLC traffics in a multi-cell network scenario. In the proposed framework, agents are deployed at the network edge to provide online executions, while a centralized offline training process is performed at the Non-RT RIC located in a regional cloud server to train deep neural networks using the collected data from all agents to overcome the slow convergence issue in DRL algorithms. *To the best knowledge of the authors, this is the first work to use the new O-RAN features to support distributed learning for eMBB-URLLC coexistence in a multi-cell wireless network.*

II. System Model and Problem Formulation

A. Network Model

Consider a multi-cell network providing two different services to users: eMBB and URLLC, as shown in Fig. 1. In the considered model, multiple edge servers are deployed at the network edge and connected to a regional cloud server. Following the O-RAN network architectures in [10], the Non-RT RIC is located at the cloud server, while Near-RT RICs are installed at the edge servers. We define $\mathcal{B} = \{1, \ldots, B\}$ as the set of all BSs, where a BS $b \in \mathcal{B}$ serves a set $\mathcal{M}_b^e = \{1, \ldots, M_b^e\}$ of eMBB users and a set $\mathcal{M}_b^u = \{1, \ldots, M_b^u\}$ of URLLC users. Furthermore, each BS is associated with one edge server. The time-frequency plan is divided into K resource blocks (RBs), where each RB spans a time interval defined as a time slot in the time domain and includes a bandwidth of f Hz in the frequency domain. Each time slot is further divided into N short transmission time intervals (sTTI), i.e., mini-slots.

Due to the stringent latency requirement of the URLLC users, we prioritize the scheduling of their arrival traffic by adopting the puncturing scheme where the arrival URLLC traffic will be scheduled immediately in the next sTTI. We define the following puncturing decision variable:

$$\alpha_{k,n}^{b,m}(t) = \begin{cases} 1, & \text{if the } n^{\text{th}} \text{ sTTI of the time sot } t \text{ is punctured} \\ & \text{by the } m^{\text{th}} \text{ URLLC user, } \forall\, k \in \mathcal{K},\ b \in \mathcal{B}, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Let $h_{b,k}^{e,m}(t)$ denote the eMBB time-varying channel gain over the k^{th} RB in the b^{th} cell, and $p_{b,k}^{e,m}(t)$ is the transmission power to the eMBB user m at the RB k. Thus, the Signal-to-Noise-and-Interference (SINR) of the eMBB user m is given by:

$$\gamma_{b,k}^{e,m}(t) = \frac{p_{b,k}^{e,m}(t) h_{b,k}^{e,m}(t)}{\sum\limits_{\substack{b' \in \mathcal{B} \\ b' \neq b}} \underbrace{p_{b',k}^{e,m}(t) h_{b',k}^{e,m}(t)}_{\text{eMBB interference}} + \sum\limits_{\substack{b' \in \mathcal{B} \\ b' \neq b}} \underbrace{p_{b',k}^{u,m}(t) h_{b',k}^{u,m}(t)}_{\text{URLLC interference}} + \sigma^2}. \tag{2}$$

Accordingly, the achieved eMBB data rate at the m^{th} user using the k^{th} channel connected to the BS b is defined as

$$r_{b,k}^{e,m}(t) = f\left(1 - \frac{\sum_{n=1}^{N} \alpha_{k,n}^{b,m}(t)}{N}\right) \log_2\left(1 + \gamma_{b,k}^{e,m}(t)\right), \tag{3}$$

the term $\frac{\sum_{n=1}^{N} \alpha_{k,n}^{b,m}(t)}{N}$ represents the eMBB data rate loss due to URLLC scheduling and the notation f represents the RB bandwidth. Let $x_{m,k}^{b}(t)$ be the RBs allocation indicator to eMBB users

$$x_{m,k}^{b}(t) = \begin{cases} 1, & \text{if RB } k \text{ is assigned to eMBB user } m, \forall b \in \\ & \mathcal{B}, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Thus, the total downlink eMBB data rate obtained by user m is given by

$$r_{b,m}^{e}(t) = \sum_{k \in \mathcal{K}} x_{m,k}^{b}(t) r_{b,k}^{e,m}(t). \tag{5}$$

The achievable URLLC rate cannot be accurately obtained using the Shannon capacity model due to the short packets nature of URLLC transmissions. Instead, the URLLC rate can be obtained in the finite blocklength regime [9]. Thus, the data rate achieved by the m^{th} URLLC user associated with the BS b at time slot t is

$$r^{u,m}_{b,k}(t) = \sum_{k\in\mathcal{K}} f_k \frac{\sum_{n=1}^{N} \alpha^{b,m}_{k,n}(t)}{N}\Bigg[\log_2\left(1+\gamma^{u,m}_{b,k}(t)\right) - \sqrt{\frac{D^{u,m}_{b,k}(t)}{c^{u,m}_{b,k}(t)}}\,Q^{-1}(\vartheta)\Bigg], \tag{6}$$

where $\gamma^{e,m}_{b,k}$ is the SINR of URLLC user m over the k^{th} RB, given by

$$\gamma^{u,m}_{b,k}(t) = \frac{p^{u,m}_{b,k}(t)h^{u,m}_{b,k}(t)}{\underbrace{\sum_{\substack{b'\in\mathcal{B}\\ b'\neq b}} p^{u,m}_{b',k}(t)h^{u,m}_{b',k}(t)}_{\text{URLLC interference}} + \underbrace{\sum_{\substack{b'\in\mathcal{B}\\ b'\neq b}} p^{e,m}_{b',k}(t)h^{e,m}_{b',k}(t)}_{\text{eMBB interference}} +\sigma^2}, \tag{7}$$

where $c^{u,m}_{b,k}(t)$ is the number symbols in each sTTI (URLLC mini-slot) called the length of the CB. The notation $Q(\cdot)$ refers to the Gaussian Q-function, and the notation $\vartheta > 0$ defines the error probability associated with URLLC transmissions. $D^{u,m}_{b,k}(t)$ is the channel dispersion, a measure of stochastic channel variations compared to a deterministic channel, defined as

$$D^{u,m}_{b,k}(t) = 1 - \frac{1}{\left(1+\gamma^{u,m}_{b,k}(t)\right)^2}. \tag{8}$$

B. Problem formulation

Scheduling URLLC traffic over eMBB users with poor channel quality causes more burden on eMBB transmissions, which may violate the minimum required QoS. Thus, we design a risk-averse objective function that includes the mean-variance tradeoff of the eMBB data rate as follows:

$$\mathcal{G}(\boldsymbol{X}, \boldsymbol{P}, \mathbf{A}) = \sum_{b\in\mathcal{B}} \sum_{m\in\mathcal{M}^e_b} \mathbb{E}_t\left[r^e_{b,m}(t)\right] - \beta\mathsf{Var}\left[r^e_{b,m}(t)\right], \tag{9}$$

where the notation $\mathbb{E}$ defines the expectation, β is a controllable weighting parameter to adjust the preference of the variance part, and Var defines the variance. In particular, the objective function (9) is designed following the Markowitz mean-variance riks-averse equation [11]. The variance term can capture the investment risk in the modern portfolio theory. Besides, in (9), the variance part characterizes the uncertainty in channel variations, which is vital in defining transmission reliability. Analogous to the investment process in the modern portfolio theory, the BSs obtain appropriate URLLC scheduling considering eMBB users with poor channel quality.

Furthermore, the transmission reliability of URLLC users is defined in terms of the transmission error probability ϑ which should not exceed a given threshold ε

$$\vartheta^u_{b,m} \leq \varepsilon, \; \forall\, m \in \mathcal{M}^u_b, \; b \in \mathcal{B}. \tag{10}$$

Let μ_m be the packet size of URLLC user m, $l_m(t)$ be the number of generated URLLC packets to user m at time slot t, and τ be the time slot duration, the data rate of the m^{th} URLLC user in can be expressed as

$$r^{u,m}_b(t) \approx \frac{\mu_m \times l_m(t)}{\tau} \tag{11}$$

Accordingly, from (10) and (11), we can define the URLLC reliability constraint as follows:

$$Q\left(\frac{\ln(1+\gamma^{u,m}_{b,k}(t)) - \frac{\mu_m \ln 2}{\tau f|\mathcal{K}^u_m|}}{\sqrt{\frac{D^u_n}{\tau f|\mathcal{K}^u_m|}}}\right) \leq \varepsilon. \tag{12}$$

Our objective is to formulate an optimization problem to optimize the eMBB RBs allocation, transmission power to eMBB users, and URLLC scheduling strategy. Therefore, we formulate the resource allocation optimization problem for eMBB and URLLC services as follows:

$$\underset{\boldsymbol{X}, \boldsymbol{P}, \mathbf{A}}{\text{maximize}} \quad \mathcal{G}(\boldsymbol{X}, \boldsymbol{P}, \mathbf{A}), \tag{13a}$$

$$\text{subject to} \quad Q\left(\frac{\ln(1+\gamma^{u,m}_{b,k}(t)) - \frac{\mu_m \ln 2}{\tau f|\mathcal{K}^u_m|}}{\sqrt{\frac{D^u_n}{\tau f|\mathcal{K}^u_m|}}}\right) \leq \varepsilon, \tag{13b}$$

$$\sum_{m\in\mathcal{M}^e_b} \sum_{k\in\mathcal{K}} p^{e,m}_{b,k}(t) \leq P_{\text{max}}, \; \forall\, b \in \mathcal{B}, \tag{13c}$$

$$\sum_{m\in\mathcal{M}^e_b} x^b_{m,k}(t) \leq 1, \; \forall\, k \in \mathcal{K}, \; b \in \mathcal{B}, \tag{13d}$$

$$\sum_{n\in\mathcal{N}} \alpha^{b,m}_{k,n}(t) \leq N, \; \forall\, k \in \mathcal{K}, \; b \in \mathcal{B}, \tag{13e}$$

$$\sum_{m\in\mathcal{M}^u_b} \alpha^{b,m}_{k,n}(t) \leq 1, \; \forall\, k \in \mathcal{K}, \; b \in \mathcal{B}, \tag{13f}$$

$$p^{e,m}_{b,k}(t) \geq 0, \; \forall\, m \in \mathcal{M}^e, \; k \in \mathcal{K}, \tag{13g}$$

$$x^b_{m,k}(t) \in \{0, 1\}, \; \forall m \in \mathcal{M}^e, \; k \in \mathcal{K}, \tag{13h}$$

$$\alpha^{b,m}_{k,n}(t) \in \{0, 1\}, \; \forall\, m \in \mathcal{M}^u, \; k \in \mathcal{K}, \tag{13i}$$

where the notation P_{max} defines the transmission power threshold of each BS. The formulated optimization problem in (13) aims to obtain the optimal resource slicing decisions that include the optimum $\boldsymbol{X}^*$, $\boldsymbol{P}^*$, and $\mathbf{A}^*$. The constraint (13b) ensures that the URLLC transmission error probability doesn't exceed a predefined value ε. Furthermore, constraints (13c), (13d), (13e), (13f), (13g), (13h), and (13i) represent the resource allocation visibility constraints. In the formulated optimization problem, we include the eMBB transmission power in the decision variables to reduce the impact of scheduling URLLC traffic over eMBB users by increasing the transmission power over users experiencing more burden. However, URLLC transmission power is set at the maximum allowed value to achieve ultra-reliable transmissions.

III. Proposed Distributed Learning Framework

Problem (13) is a mixed-integer programming that is NP-hard in general. Moreover, the URLLC scheduling variable is coupled with the RB and power allocation variables, increasing the complexity of the optimization problem. In particular, an online solution is essential to satisfy the sensitive delay

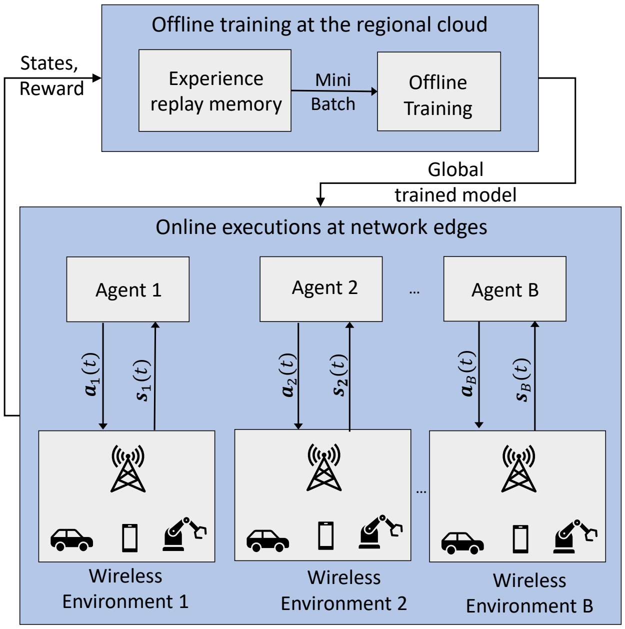

Figure 2: Proposed distributed learning framework.

requirement of URLLC users and cope with the high network dynamic caused by users' mobility. However, it is hard to obtain an online solution using the typical optimization methods, such as convex optimization, as (13) is an NP-hard problem. Furthermore, the randomized nature of URLLC traffic necessitates the need for a dynamic resource allocation technique. These challenges drive us to use DRL techniques to solve the formulated optimization problem.

A. Markov Model for the Multi-Cell Cooperation Network

We transform the formulated problem (13) into a Markov game for B agents. The Markov game is defined as a set of states $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_b, \cdots, \mathcal{S}_B\}$, and a set of actions $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_b, \cdots, \mathcal{A}_B\}$. For a given state $s_b(t) \in \mathcal{S}_b$, the corresponding agent uses the policies, $\pi_b : \mathcal{S}_b \mapsto \mathcal{A}_b$ to choose an action from their action spaces according to their observations corresponding to $s_b(t)$.

The state space contains the channel states of all eMBB and URLLC users and the network traffic status at each decision step, i.e., time slot. For instance, the state of an agent b at time slot t is a vector $\boldsymbol{s}_b(t) = \{\boldsymbol{h^e}_b(t), \boldsymbol{h^u_b(t)}, \lambda_b, M^e_b, M^u_b\}$. Each edge server[2] is regarded as an agent. We consider that each agent receives only its own state, i.e., users' information in the same cell, to reduce the overheads caused by information exchange among the cells. Each agent has an actor-network that decides the agent's decision, i.e., the action selection policy. The actions of each agent are the output of its actor-network which contains the decision variables of the optimization problem (13). Thus, the action space is defined as $\mathcal{A} = \{\boldsymbol{X}, \boldsymbol{P}, \mathbf{A}\}$.

Let $R(t) : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$ be the instantaneous returned reward at time slot t. The requirements of each service should be considered in designing the reward function. Thus, the reward function is formulated in the following two parts:

$$R^e(t) = \sum_{b \in \mathcal{B}} \sum_{m \in \mathcal{M}^e_b} \mathbb{E}_t\big[r^e_{b,m}(t)\big] - \beta \mathsf{Var}\big[r^e_{b,m}(t)\big], \tag{14}$$

$$R^u(t+1) = Q\left(\frac{\ln(1+\gamma^{u,m}_{b,k}(t)) - \frac{\mu_m \ln 2}{\tau f |\mathcal{K}^u_m|}}{\sqrt{\frac{D^u_n}{\tau f |\mathcal{K}^u_m|}}}\right) - \varepsilon, \tag{15}$$

where $R^e(t)$ defines the data rate and reliability requirements of eMBB users and $R^u(t)$ captures the URLLC reliability. Accordingly, the total reward function is given by

$$R(t) = R^e(t) - \phi(t) R^u(t), \tag{16}$$

where $\phi(t)$ is a weighting parameter that varies over time slots to satisfy the required URLLC reliability. Here, we update the values of the parameter $\phi(t)$ over time slots as follows:

$$\phi(t+1) = \max\left\{\phi(t) + \vartheta(t) - \varepsilon, 0\right\}, \tag{17}$$

where $\vartheta(t)$ is the obtained transmission error probability at the t^{th} time slot. In particular, adjusting the value of $\phi(t)$ over time slots allows us to verify URLLC reliability dynamically per service requirements.

B. Multi-Agent DRL Algorithm

As depicted in Fig. 2, the DRL agents are distributed at the network edges, whereas a central training unit is located at the regional cloud server to ease implementation and improve stability. The centralized server trains a global model using the gathered experiences from all edge agents. This approach allows agents to learn together for faster convergence and better performance. Furthermore, the decision made by each agent is unaware to others. Here, sharing the same learning parameters by the central trainer to all agents still gives different action decisions by the agents as they execute the trained model with different local states.

A fully connected neural networks model is trained offline at the Non-RT RIC, installed at the regional cloud, using the collected data from all edge agents. The trained model is then signaled to the DRL agents at Near-RT RICs installed at the edge servers. The global model is trained with an objective to maximize the designed global reward function in (16). Specifically, a policy gradient-based learning algorithm is adopted in Actor-Critic networks. The actor part makes decisions on action selection according to the learned policy π, while the critic network evaluates the decided actions. We use the experience replay technology with a buffer size of D to improve the stabilization of the training process. The central training unit samples a mini-batch with size d from the replay buffer to train the actor-critic networks.

The state-action function is given by

$$Q\left(s, a\right) = \sum_{t=0}^{\infty} \zeta R(t+1) \mid \boldsymbol{\pi}, s = s(t), a = a(t), \tag{18}$$

[2] In the considered network model, each edge server is associated with one BS.

where ζ is a discount factor. The network objective function $J(\pi)$ is defined as

$$J(\pi) = \mathbb{E}\Big[Q^{\pi}(\boldsymbol{s},\boldsymbol{a})\Big] = \int_S \int_A \pi(\boldsymbol{s},\boldsymbol{a}) Q^{\pi}(\boldsymbol{s},\boldsymbol{a}) dads. \quad (19)$$

At the actor part, the policy is initialized according to the network parameter $\boldsymbol{\theta}$ as follows:

$$\pi(\boldsymbol{s},\boldsymbol{a};\boldsymbol{\theta}) = Pr(\boldsymbol{a}|\boldsymbol{s},\boldsymbol{\theta}). \quad (20)$$

We obtain the objective function gradient with respect to $\boldsymbol{\theta}$ in the following equation:

$$\nabla_{\boldsymbol{\theta}} J(\pi) = \int_S \int_A \nabla_{\pi} Q^{\pi}(\boldsymbol{s},\boldsymbol{a}) dads. \quad (21)$$

Then, the actor network paraemnters $\boldsymbol{\theta}$ is updated based on the (21) as

$$\boldsymbol{\theta}(t+1) = \boldsymbol{\theta}(t) + \rho_a \nabla_{\boldsymbol{\theta}} J(\pi), \quad (22)$$

where ρ_a represents the actor network's learning rate.

At the critic unit, we use the function estimator technique to obtain an approximation of the state-action function $Q^{\pi}(\boldsymbol{s},\boldsymbol{a})$. Thus, the approximated state-action value function using the linear function estimator is give by

$$V(\boldsymbol{s},\boldsymbol{a}) = \boldsymbol{\xi}^T \boldsymbol{\kappa}(\boldsymbol{s},\boldsymbol{a}) = \sum_{i\in\mathcal{S}} \xi_i \kappa_i(\boldsymbol{s},\boldsymbol{a}), \quad (23)$$

where $\boldsymbol{\kappa} = [\kappa_1(\boldsymbol{s},\boldsymbol{a}),\ldots,\kappa_S(\boldsymbol{s},\boldsymbol{a})]^T$ is a basis function, $\boldsymbol{\xi}(\boldsymbol{s},\boldsymbol{a}) = (\xi_1,\ldots,\xi_S)^T$ is a weight vector. The Temporal-Difference (TD) technique is used to find the error in the estimated values as compared to the real values

$$\delta(t) = R(t+1) + \zeta V(t+1) - V(t). \quad (24)$$

We leverage the gradient descent technique to update the weighting vector $\boldsymbol{\xi}(\boldsymbol{s},\boldsymbol{a})$ as follows:

$$\boldsymbol{\xi}(t+1) = \boldsymbol{\xi}(t) + \rho_c \delta(t) \nabla_{\xi} V(\boldsymbol{s},\boldsymbol{a}), \quad (25)$$

where ρ_c denotes the learning rate of the critic network. Finally, the value function in (23) is updated according to value of $\boldsymbol{\xi}(\boldsymbol{s},\boldsymbol{a})$ in (25).

At the edge servers, each agent uses the received trained model from the Non-RT RIC to determine the best resource allocation policy based on the given local network states. The agents' reward is obtained according to the selected policy and network states. The edge agents send the network observations and the obtained reward to the experience replay memory at the regional cloud server for improving the trained models over time.

IV. Performance Evaluation

A wireless network composed of three BSs is considered. Each BS covers an area of 200 m^2 and serves eMBB and URLLC users with a constant full-buffer traffic rate and Poisson traffic with arrival rate (λ), respectively. URLLC packets length is set to be 125 Bytes [9]. Moreover, users' mobility is considered with time-varying path loss and channel conditions. The path loss between users and the associated BS is given as $120.8 + 37.5\log_{10}(d)$ dB, where d defines the BS-user distance in meters. The AWGN is set as $\sigma^2 = -114$ dBm. The maximum transmission power threshold is 38 dBm. The period of each time slot is configured at 1 ms and contains 7 mini-slots. The bandwidth of each RBs is set at 180 kHz [7]. We train the proposed algorithm under different network settings, e.g., different URLLC traffic rates, varying the distance between BSs and users. A neural network model consisting of three hidden layers with 600, 300, and 250 neurons, respectively, is used. The discount factor is set at 0.95, while the learning rate is adjusted at 0.001. The reply buffer memory size is set to 2000, and the mini-batch size is configured at 32 [9].

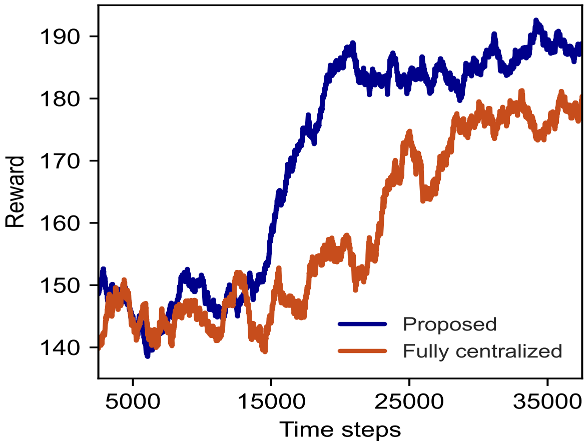

Figure 3: Convergence of the reward function.

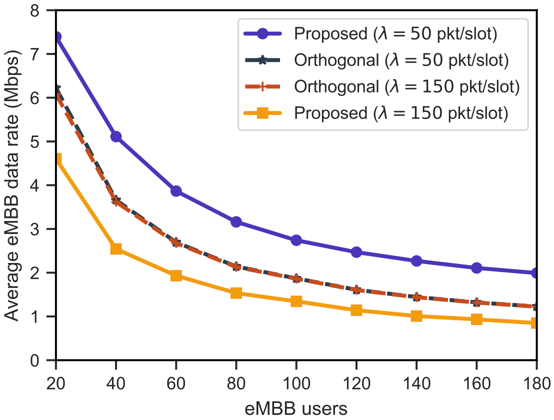

Figure 4: Average eMBB data rate for different network size settings.

Algorithm Convergence: We discuss the convergence rate of the proposed algorithm and compare it to the *Fully Centralized* approach, where a central agent, meta-agent, is trained using states collected from all agents. In such a scenario, the meta-agent decides the action selection of all agents and then forwards the results to BSs. As depicted in Fig. 3, the *Fully Centralized* approach incurs a worse performance and requires a longer convergence time. This is due that the action space of the meta-agent contains the joint actions of all BSs, increasing the dimension of the action space. On the other hand, the

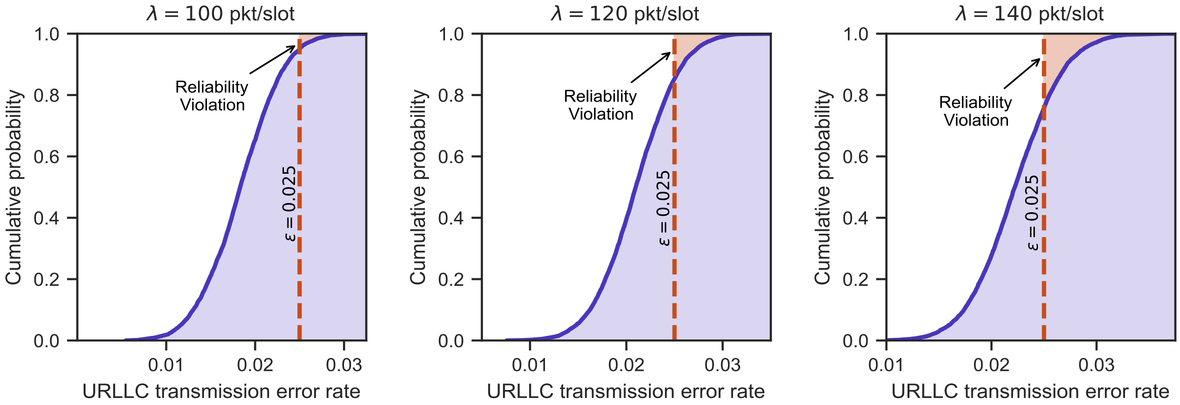

Figure 5: CDF of URLLC transmission error probability for different traffic settings.

proposed approach has a fast convergence rate, achieving a better response to the dynamic environment.

eMBB Data Rate: We compare the proposed method to the *orthogonal* resource allocation technique in terms of the obtained average eMBB data rate in Fig. 4. The results show that the proposed approach performs better under light URLLC traffic ($\lambda = 50$ packets per time slot). However, the performance of the proposed algorithm is lower than the orthogonal method when increasing the URLLC traffic rate. In particular, the proposed algorithm schedules URLLC traffic over the ongoing eMBB transmissions, giving higher priority to URLLC service. This impacts eMBB transmissions under a heavy URLLC traffic rate scenario. It is also noticeable that the *orthogonal* method gives the same performance under different URLLC traffic settings as this approach allocates fixed resources to each service regardless of the URLLC traffic. Moreover, Fig. 4 shows that the average eMBB data rate reduces when the number of eMBB users is increased for the same available resources.

URLLC Reliability: Finally, we discuss the URLLC transmission error rate obtained by the proposed algorithm in Fig. 5. Here, URLLC reliability is defined in terms of the violation probability of (12). To do so, we plot the Cumulative Distribution Function (CDF) of the transmission error probability of URLLC traffic for different settings of λ at $\varepsilon = 0.025$ to emphasize the worst-case scenario. It is noticeable that the transmission error rate falls lower than the predefined threshold ε with a probability higher than 0.99 when setting $\lambda = 100$ packet/time slot. In fact, the proposed URLLC scheduling method allocates resources to URLLC users by learning the network traffic and channel variations, which improve transmission reliability. It is also shown in Fig. 5 that increasing URLLC traffic rate over the same network resources may cause a violation of URLLC reliability. This is because the proposed algorithm schedules URLLC traffic over eMBB transmissions considering the trade-off between eMBB and URLLC requirements.

V. Conclusion

This paper has studied the resource allocation problem in multi-cell wireless systems serving two types of users, eMBB and URLLC. We first formulated a risk-averse optimization problem that incorporates the requirements of each traffic type. A distributed learning framework has been developed considering the novel O-RAN network architectures that facilitate learning over wireless networks to solve the resource allocation problem. In particular, a multi-agent DRL-based algorithm has been developed that can provide online decisions on resource allocation by deploying trained execution agents at Near-RT RICs located at network edges. Simulation results have shown the performance of the proposed algorithm in satisfying the QoS requirements of both eMBB and URLLC users.

References

[1] O.-R. W. 1, "O-ran architecture description — v2.00," Tech. Rep., 2020.

[2] B. Balasubramanian, E. S. Daniels, M. Hiltunen, R. Jana, K. Joshi, R. Sivaraj, T. X. Tran, and C. Wang, "Ric: A ran intelligent controller platform for ai-enabled cellular networks," *IEEE Internet Computing*, vol. 25, no. 2, pp. 7–17, 2021.

[3] 3GPP, "Technical specification group services and system aspects; release 15 description," Tech. Rep., TR 21.915, v1.1.0, March 2019.

[4] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5g wireless network slicing for embb, urllc, and mmtc: A communication-theoretic view," *IEEE Access*, vol. 6, pp. 55 765–55 779, 2018.

[5] A. Anand, G. D. Veciana, and S. Shakkottai, "Joint scheduling of URLLC and eMBB traffic in 5g wireless networks," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*. IEEE, apr 2018.

[6] J. Park and M. Bennis, "URLLC-eMBB slicing to support VR multimodal perceptions over wireless cellular systems," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, dec 2018.

[7] A. K. Bairagi, M. S. Munir, M. Alsenwi, N. H. Tran, S. S. Alshamrani, M. Masud, Z. Han, and C. S. Hong, "Coexistence mechanism between embb and urllc in 5g wireless networks," *IEEE Transactions on Communications*, vol. 69, no. 3, pp. 1736–1749, 2021.

[8] P. Korrai, E. Lagunas, S. K. Sharma, S. Chatzinotas, A. Bandi, and B. Ottersten, "A ran resource slicing mechanism for multiplexing of embb and urllc services in ofdma based 5g wireless networks," *IEEE Access*, vol. 8, pp. 45 674–45 688, 2020.

[9] M. Alsenwi, N. H. Tran, M. Bennis, S. R. Pandey, A. K. Bairagi, and C. S. Hong, "Intelligent resource slicing for embb and urllc coexistence in 5g and beyond: A deep reinforcement learning based approach," *IEEE Transactions on Wireless Communications*, vol. 20, no. 7, pp. 4585–4600, 2021.

[10] O. Alliance, "O-ran use cases and deployment scenarios," *White Paper*, 2020.

[11] H. M. Markowitz and G. P. Todd, *Mean-variance analysis in portfolio choice and capital markets*. John Wiley & Sons, 2000, vol. 66.