Dynamic Resource Slicing of eMBB/URLLC Traffics in 5G Wireless Networks: A Reinforcement Learning Based Approach

Madyan Alsenwi, Shashi Raj Pandey, Yan Kyaw Tun, and Choong Seon Hong
Department of Computer Science and Engineering, Kyung Hee University, South Korea
{malsenwi, shashiraj, ykyawtun7, cshong}@khu.ac.kr

Abstract

Ultra Reliable Low Latency Communications (URLLC) is a 5G wireless networks service that accommodates applications having strict reliability and latency requirements such as industrial IoT, virtual reality, and autonomous vehicles. URLLC aims low latency and high reliability transmissions. In this paper, the dynamic multiplexing of URLLC and enhanced Mobile Broad Band (eMBB) is tackled. We model the problem as a Markov Decision Process. Moreover, We propose a weighted formulation that takes into account both the total average data rate of eMBB traffic and the URLLC reliability. To this end, We propose a Q-learning algorithm where the gNB learns the optimal resource allocation to eMBB and URLLC traffics. The simulation results illustrate the performance of the proposed algorithm.

Keywords - 5G New Radio, eMBB, URLLC, Resource Slicing, Q-learning.

1. Introduction

5G New Radio (NR) supports different services with different requirements (i.e., enhanced Mobile Broad Band (eMBB), and URLLC). While eMBB is designed to maximize the system data rate which is an extension of LTE-Advanced, URLLC is designed to support applications that require ultra-high reliability and low latency. As mentioned in the Third Generation Partnership Project (3GPP), URLLC provides services with a packet error rate less than 10^{-5} while ensuring that the latency is less than $0.5ms$ which are very critical in some applications such as autonomous vehicle, and industrial IoT [1].

Supporting these services with different requirements increases the complexity of resource slicing in 5G NR. One approach to distribute resources among theses services is the orthogonal-slicing. In this technique, slices are allocated to each traffic orthogonality (i.e., slices are orthogonal in frequency-time) considering the requirements of each service. The orthogonal-slicing gives a free-interference resource slicing. However, This approach leads to inefficient use of available resources [2]. Another approach, called non-orthogonal slicing, is to share resources between these different services and allocate resources dynamically. Here, reusing resources among the different services improves the spectral efficiency. However, the non-orthogonal slicing may cause interference among these different services [3, 4].

In this work, we consider the URLLC/eMBB resource slicing problem for non-orthogonal approach. Here, the time domain is divided into equally spaced time slots (i.e. eMBB transmit time interval). Each time slot is further divided into mini-slots (i.e. URLLC transmit time interval) [1]. The incoming URLLC is scheduled immediately in the next mini-slot to achieve a low latency transmission. The available resources are allocated dynamically to eMBB/URLLC traffics considering the eMBB user's data rate and the URLLC reliability. We formulate the problem of resource allocation as a Markov Decision Problem (MDP) wherein the state represents the minislots, the decision variable is the number of RBs allocated to URLLC traffic. The total expected reward is defined as a weighted summation of total eMBB data rate and URLLC reliability. The URLLC reliability is formulated based on the outage probability of URLLC traffic.

2. System Model

We consider a gNB with M eMBB users. The total resources are divided into a number of Resource Blocks (RBs). The RBs allocated to eMBB transmission at the slots level (i.e., at the beginning of each tiem slot). However, the arrived URLLC traffic during a time slot cannot be delayed to the next time slot due to its latencey requirement. Therefore, we schedule the incoming URLLC traffic at the level of minislots as shown in Fig. 1. Hence, scheduling the incoming URLLC traffic at the edge of minislot ensures its latency constraint. However, scheduling the URLLC traffic over minislots of a time slot that already assigned to eMBB traffic causes interference to eMBB transmission. Hence, the gNB stops the eMBB transmissions (i.e., by allocating zero power to it) overlapped with URLLC (i.e., puncturing eMBB transmission) [5, 6, 7].

Puncturing eMBB transmissions impacts the data rate of eMBB users. Therefore, this work aims to minimize the impacting on eMBB data rate while considering the URLLC requirements. Here, we approximate the loss of eMBB data rate as a linear proportional to the punctured RBs by the URLLC. Therefore, the eMBB users data rate can be approximated as follows:

$$R_{eMBB} = \sum_{m=1}^{M} (b_m - l_m) f \log_2(1 + SNR_m), \quad (1)$$

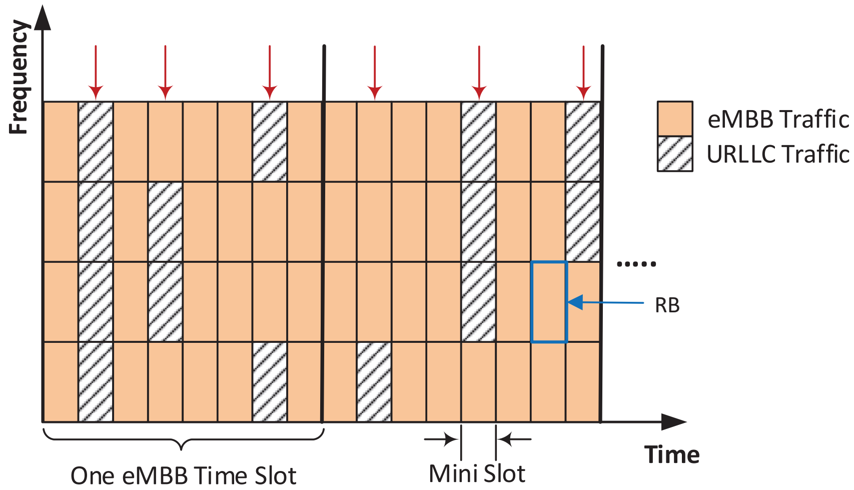

Figure 1: Puncturing approach for multiplexing of URLLC and eMBB

where b_m is the number of RBs allocated to the m^{th} eMBB user, l_m is the number of punctured RBs from the m^{th} eMBB user due to URLLC traffic, f is the bandwidth of each RB (i.e., we consider that all RBs have the same bandwidth), SNR_m is the eMBB user m signal to noise ratio, and M is the total number of eMBB users.

We consider that Z is a random variable representing the URLLC traffic load. Therefore, we can calculate the outage probability of URLLC traffic as follows:

$$P_{outage} = P[\sum_{m=1}^{M} l_m f \log_2(1 + SNR_u) < Z] \leq \eta, \quad (2)$$

Where η is the maximum outage probability, and SNR_u is the signal to noise ratio of URLLC.

The outage probability can be transformed into a deterministic form using the Cumulative Distribution Function (CDF) of Z. Therefore, we can write the outage probability as follows:

$$\begin{aligned} P_{outage} &= P[Z > \sum_{m=1}^{M} l_m f \log_2(1 + SNR_u)] \leq \eta \\ &= 1 - F_Z(\sum_{m=1}^{M} l_m f \log_2(1 + SNR_u)) \leq \eta \quad (3) \\ &= \sum_{m=1}^{M} l_m f \log_2(1 + SNR_u) \geq F_Z^{-1}(1 - \eta) \end{aligned}$$

Where $F_Z(x)$ is the CDF of the random variable Z.

Here, we consider that the URLLC traffic follows the Pareto Distribution (i.e., the random variable Z follows the Pareto Distribution). Therefore, the CDF of Z is defined as:

$$F_Z(z) = P(Z \leq z) = 1 - (\frac{z_n}{z})^{\lambda}, \quad for\ z \geq z_n, \quad (4)$$

Where z_n is the minimum value of Z, and λ takes any positive number.

3. Proposed Reinforcement Learning Algorithm

Here, We address the dynamic multiplexing of URLLC/eMBB traffics problem within a MDP. We define the following MDP:

- *States:* each state $x^{(t)} \in \mathcal{X}$ corresponds to a minislot in time domain as shown in Fig. 2.
- *Action:* The gNB autonomously responds to the URLLC traffic and network state based on the historical data and selects an action from the set $\mathcal{A} = \{v_0, v_1, ..., v_B\}$ where B is the total number of RBs, and v_b is the number of RBs allocated to URLLC traffic which takes a value in the range of $[0, B]$.
- *Reward:* We define the reward function $R(x^{(t)}, l)$ as a weighted summation of eMBB data rate and the URLLC reliability. Therefore, the reward function can be given as follows:

$$\begin{aligned} R(x^{(t)}, l) = &\sum_{m=1}^{M} (b_m - l_m) f \log_2(1 + SNR_m) \\ &+\beta\Big((\sum_{m=1}^{M} l_m f \log_2(1 + SNR_u))^{\lambda} - \frac{x_n^{\lambda}}{\eta}\Big) \end{aligned} \quad (5)$$

The gNB estimates the value of each state-action pair. It updates its estimates through the interaction with the environment where at each iteration it performs an action and then observes the reward. The gNB chooses an action at each minislot following the ϵ-greedy policy, that is, it selects

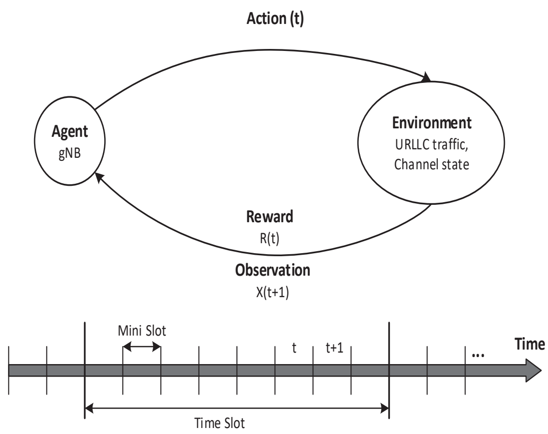

Figure 2: Reinforcemetn Learning Model

an action that maximizes its current estimate with probability $1-\epsilon$, or a random action with probability ϵ. The parameter ϵ captures the exploration/exploitation trade-off.

The Q-value of current state is given as [8]:

$$Q(x^{(t)}, v^{(t)}) \leftarrow (1-\alpha)Q(x^{(t)}, v^{(t)}) + \alpha\Big(R + \gamma V\big(x^{(t+1)}\big)\Big) \quad (6)$$

where α is the learning rate parameter, γ is the discount factor, and $V(x^{(t)})$ is defined as

$$V(x^{(t)}) = max_v Q(x^{(t)}, v) \quad (7)$$

4. Performance Evaluation

Here, we study the performance of the proposed model. We consider a s system with a total bandwidth of 100 RBs. The URLLC traffic is generated from Pareto Distribution with $\lambda = 2$. The agent training parameters are $\alpha = 0.9$, $\gamma = 0.9$, and $\epsilon = 0.9$.

Fig. 3 shows the convergence of the total reward under successive explorations following the Q-learning approach. As shown in Fig. 3, the reward value increases with number of episodes and this means that the proposed algorithm allocates resources to eMBB and URLLC traffics while satisfying both the eMBB data rate and URLLC reliability.

5. Conclusion

In this work, we studied the dynamic multiplexing of URLLC and eMBB transmissions. The dynamic multiplexing problem of eMBB/URLLC has been formulated as a MDP. We have formulated the reward function as a weighted summation of eMBB data rate and URLLC reliability. We introduced the Q-learning algorithm where the gNB learns the optimal resource allocation to eMBB and URLLC traffics. The results show that the the proposed algorithm can

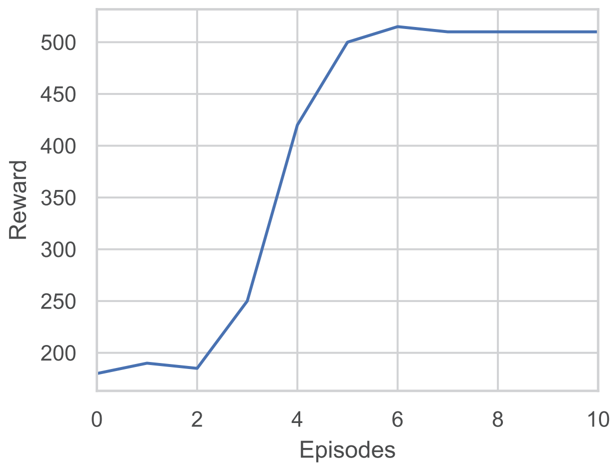

Figure 3: Rewards versus Episodes

allocate resources dynamically to eMBB and URLLC traffics efficiently.

Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (NRF-2017R1A2A2A05000995). *Dr. CS Hong is the corresponding author.

References

[1] "Ul inter-ue multiplexing between embb and urllc, document 3gpp tsg ran wg1 95, nov. 2018." Tech. Rep.

[2] S. E. Elayoubi, P. Brown, M. Deghel, and A. Galindo-Serrano, "Radio resource allocation and retransmission schemes for URLLC over 5g networks," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 4, pp. 896–904, apr 2019.

[3] M. Alsenwi, N. H. Tran, M. Bennis, A. K. Bairagi, and C. S. Hong, "eMBB-URLLC resource slicing: A risk-sensitive approach," *IEEE Communications Letters*, vol. 23, no. 4, pp. 740–743, apr 2019.

[4] S. R. Pandey, M. Alsenwi, Y. K. Tun, and C. S. Hong, "A downlink resource scheduling strategy for URLLC traffic," in *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE, feb 2019.

[5] M. Alsenwi and C. S. Hong, "Resource scheduling of urllc/embb traffics in 5g new radio," *Korean Computer Congress (KCC), South Korea*, pp. 1271–1273, 2018.

[6] A. K. Bairagi, M. S. Munir, M. Alsenwi, N. H. Tran, and C. S. Hong, "A matching based coexistence mechanism between embb and uRLLC in 5G wireless networks," in *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*. ACM, 2019, pp. 2377–2384.

[7] M. Alsenwi, S. R. Pandey, Y. K. Tun, K. T. Kim, and C. S. Hong, "A chance constrained based formulation for dynamic multiplexing of embb-urllc traffics in 5g new radio," in *The International Conference on Information Networking (ICOIN 2019), Jan. 9-11, 2019, Kuala Lumpur, Malaysia.*

[8] S. R. Pandey, S. Suhail, S. I. Moon, and C. S. Hong, "Q-learning supplemented response based crowdsensing framework for resource constrained devices," *KIISE Transaction on Computing Practices, Vol. 24, No. 7, pp. 345-351, 2018.7.*