Joint Resource Scheduling for Coexistence of URLLC and eMBB in 5G Wireless Networks

Haipeng Sun[1], Jin Yang[2*], Junhao Su[1], Haiyang Wang[1], Danpu Liu[2]
[1]Shandong Electric Power Engineering Consulting Institute Co. Ltd. , Jinan, China;
[2]School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China
email: sunhaipeng@sdepci.com, jin.yang@bupt.edu.cn, sujunhao@sdepci.com, wanghaiyang@sdepci.com, dpliu@bupt.edu.cn

2021 Computing, Communications and IoT Applications (ComComAp) | 978-1-6654-2798-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/COMCOMAP53641.2021.9653121

***Abstract*—To enable the heterogenous supports of Ultra-Reliable Low Latency Communications (URLLC) and enhanced Mobile Broadband (eMBB) traffic in 5G wireless networks, the puncturing framework is utilized to satisfy the latency and reliability requirements imposed by URLLC. However, the eMBB performance of transmission rate is degraded owing to the punctures. In this paper, we propose an effective joint resource scheduling scheme for coexistence of URLLC and eMBB traffic with the objective of minimizing the eMBB rate loss. Specifically, a more practical eMBB rate loss convex model is adopted in view of the certain error correction ability of eMBB users, and not only channel state but URLLC traffic arrival characteristics are taken into account in eMBB resource allocation. The simulation results demonstrate that the proposed scheme achieves effective gain of eMBB rate over other baseline approaches.**

***Index Terms*—URLLC, eMBB, coexistence, resource allocation, puncture.**

I. Introduction

International Telecommunication Union (ITU) defines three application scenarios based on the diversified QoS requirements on 5G wireless networks, which are Ultra-Reliable Low Latency Communications (URLLC), enhanced Mobile Broadband (eMBB) and massive Machine Type Communication (mMTC) [1]. Among them, eMBB is an extension of the existing mobile broadband scene [2] aiming at maximizing the data rate of stable connections with a moderate reliability [3]. Unlike eMBB, URLLC is designed to meet end-to-end latency requirements as low as 0.25∼0.30ms/packet while ensuring ultra-high reliability up to 99.999% [4], [5]. Hence, the high priority resource scheduling strategy needs to be adopted in order to ensure the instant transmission considering the intermittent characteristic of URLLC traffic [6].

In order to meet the heterogenous supporting requirements with current wireless network model [7], various studies on multiplexing of URLLC and eMBB traffic based on innovative superposition or puncturing framework are gaining increased attention [8], [9]. Generally, eMBB has a high occupancy rate of spectrum resources as the main bearer service [10]. To support the strict latency demands of URLLC, superposition scheme allows both URLLC and eMBB traffic to be transmitted over the same spectrum utilizing different transmission powers [11]-[13]. To reduce the intervention of URLLC on eMBB performance, the study in [14] proposed the contract-based superposition framework via matching two kinds of users with information of their distance and channel condition. However, on account of the URLLC receivers' decoding performance limitation [15], superposition is not a practical scheduling scheme in current wireless cellular networks [16].

Puncturing framework is a more appropriate method ensuring both latency and reliability requirements imposed by URLLC, where certain parts of eMBB resources are selected to load URLLC traffic, and eMBB transmits with zero power over the resources preempted by URLLC. These occupied resources referred to as punctures would result in the rate loss of eMBB traffic. To find the proper punctures placement that can reduce eMBB rate loss as much as possible, authors in [17]-[19] adopted the methods of machine learning to obtain efficient tradeoff performance under puncturing framework. In the recent work of [20], a penalty successive upper bound minimization (PSUM) scheme is proposed to resolve scheduling issue of joint URLLC and eMBB traffic. Scheduling of a non-interference subspace traffic preemption is provided in [21] for the user-centered networks. [22] considered a joint resource allocation mechanism based on the achieved eMBB rate and URLLC placement strategy to selectively overlap the ongoing eMBB transmission. In short, the aforementioned studies have paid the most attention on the punctures placement strategy on condition of given eMBB resource allocation. However, the rate loss depends on both URLLC traffic placement and eMBB resource allocation. That means it is possible to obtain additional gain through the joint scheduling of eMBB and URLLC traffic. Besides, the linear rate loss model adopted in most studies does not conform to the actual process of decoding for ignoring the error correction ability of eMBB users.

In this paper, we propose an effective joint resource scheduling scheme in the scenario of the incumbent eMBB users selectively punctured by URLLC traffic. Specifically, a more practical eMBB rate loss convex model is adopted instead of the linear model in view of the certain error correction ability of eMBB users. Then we formulate an optimization problem of jointly scheduling eMBB and URLLC traffic, and propose a heuristic scheme consisting of two steps to solve the problem. Accordingly, both channel state and URLLC traffic arrival characteristics are considered in eMBB resource allocation.

The simulation results demonstrate that the proposed scheme achieved an effective gain of eMBB rate over other baseline approaches.

The remainder of the paper is systematized as follows. Section II outlines the system model and problem formulation. Subsequently, the proposed scheme of the above-mentioned problem is addressed in Section III. Simulation and analysis are given in Section IV. Finally, conclusion is presented in Section V.

II. System Model and Problem Formulation

A. System Model

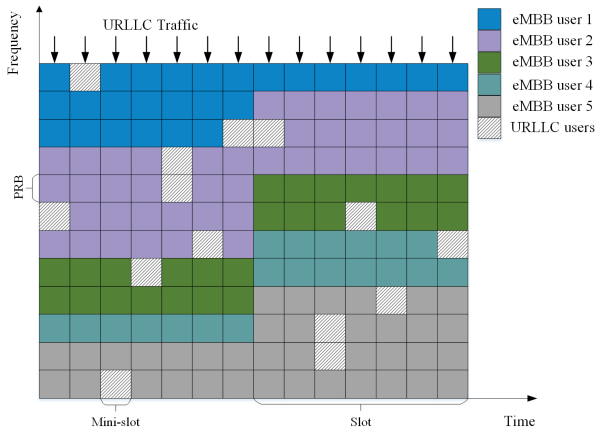

Fig. 1. Puncturing framework for coexistence of eMBB and URLLC.

In the system model, we consider a scenario of downlink transmissions with a next generation base station (gNB) supporting a set of eMBB users $\mathcal{E} = \{1, \dots |\mathcal{E}|\}$ and URLLC traffic.As depicted in Fig. 1, the available bandwidth is divided into physical resource blocks (PRBs) $\mathcal{B}$, and each PRB $b \in \mathcal{B} = \{1, \dots |\mathcal{B}|\}$ consists of 12 sub-carriers. In the transmission duration $\mathcal{T}$, the scheduling interval of the eMBB service is a time slot composed of 14 orthogonal frequency division multiplexing (OFDM) symbols. Meanwhile the URLLC scheduling interval is sliced into a more elaborate time scale of 2 OFDM symbols, which is referred to as mini-slot denoted by $m \in \mathcal{M} = \{1, \dots |\mathcal{M}|\}$ in one slot. When the system preempts the time-frequency resources of the eMBB users to load the intermittent URLLC traffic, the transmission rate of eMBB users will be down to some extent. The channel state $\mathcal{S} = \{1, \dots |\mathcal{S}|\}$ of wireless system is characterized by the $|\mathcal{E}| \times |\mathcal{S}|$ peak rate matrix $\mathcal{R}$ with the component $\widehat{r_e^s}(t)$ denoting the peak rate of eMBB user $e \in \mathcal{E}$ in slot t in channal state s.

B. Problem Formulation

As the definition of URLLC traffic in the 3rd Generation Partnership Project (3GPP) is ftp-3 model, which is a Poisson process with arrival rate λ. let $D(t)$ denote arrival URLLC traffic demand in slot t. As far as the rate loss of eMBB users is concerned, the total punctured amount is what actually counts instead of the specific punctured mini-slots and PRBs. Hence, let the variable $L_{e,m}^{\pi,s}(t)$ represent the URLLC traffic load when the user $e \in \mathcal{E}$ adopts the scheduling scheme π at mini-slot m in channel state s. The URLLC traffic load of user e in slot t is denoted by variable $L_e^{\pi,s}(t)$. According to the relationship between URLLC arrival traffic and loads of eMBB users, we have:

$$E\left[D\left(t\right)\right] = \sum_{e \in \mathcal{E}} L_e^{\pi,s}\left(t\right) = \sum_{e \in \mathcal{E}} \sum_{m \in \mathcal{M}} L_{e,m}^{\pi,s}\left(t\right) \tag{1}$$

There are three rate loss models described mathematically in [23], which respectively are linear, convex and threshold models. Considering eMBB users have certain error correction ability in the actual decoding process, the relationship between the rate loss and the URLLC load is not directly linear. Further, the simulation in [18] demonstrate the influence of punctures on the block error rate is convex. Hence a more realistic convex function $f_e^{\pi,s}\left(\cdot\right)$ is adopted to represent the eMBB transmission rate loss:

$$f_e^{\pi,s}\left(x\right) = \begin{cases} \left(\dfrac{x}{0.9}\right)^2, & \text{if} \quad x \leq 0.9, \\ 1, & \text{if} \quad 0.9 < x \leq 1. \end{cases} \tag{2}$$

The transmission rate of eMBB user e under scheduling scheme π at slot t in channel state s is given by:

$$r_e^{\pi,s}\left(t\right) = \sum_{b \in \mathcal{B}} x_{e,b}^{\pi,s}\left(t\right) \widehat{r_e^s}(t) \left[1 - f_e^{\pi,s}\left(\frac{L_e^{\pi,s}\left(t\right)}{|M| \sum\limits_{b \in \mathcal{B}} x_{e,b}^{\pi,s}\left(t\right)}\right)\right] \tag{3}$$

where $f_e^{\pi,s}\left(\cdot\right)$ is the eMBB transmission rate loss function. The variable $x_{e,b}^{\pi,s}\left(t\right)$ indicates whether the PRB b in slot t is allocated to the eMBB user e under scheduling scheme π when the system is in the channel state s.

$$x_{e,b}^{\pi,s}\left(t\right) = \begin{cases} 1, \text{ if the PRB } b \text{ is allocated to user } e \text{ in slot } t, \\ 0, \text{ otherwise.} \end{cases} \tag{4}$$

To reduce the degradation of eMBB performance due to URLLC preemption, the objective in most studies is to maximize the sum rate of all eMBB users. However, all these studies only attempt to find the $L_{e,m}^{\pi,s}\left(t\right)$ on condition of given eMBB resource allocation, although the eMBB rate depends on both $L_{e,m}^{\pi,s}\left(t\right)$ and $x_{e,b}^{\pi,s}\left(t\right)$ as shown in Eq. (3). Therefore, we aim to optimize both URLLC placement $L_{e,m}^{\pi,s}\left(t\right)$ and eMBB resource allocation $x_{e,b}^{\pi,s}\left(t\right)$ in this paper. The optimization problem can be formulated as follows:

$$\max_{x^{\pi}, L^{\pi}} \sum_{e \in \mathcal{E}} \sum_{b \in \mathcal{B}} x_{e,b}^{\pi,s}\left(t\right) \widehat{r_e^s}(t) \left[1 - f_e^{\pi,s}\left(\frac{L_e^{\pi,s}\left(t\right)}{|M| \sum\limits_{b \in \mathcal{B}} x_{e,b}^{\pi,s}\left(t\right)}\right)\right] \tag{5}$$

$$s.t. \quad x_{e,b}^{\pi,s}\left(t\right) \in \{0,1\}, \ \forall e \in \mathcal{E}, \ b \in \mathcal{B}, \tag{6}$$

$$\sum_{e \in \mathcal{E}} x_{e,b}^{\pi,s}\left(t\right) \leq 1, \ \forall b \in \mathcal{B}, \tag{7}$$

$$\sum_{b\in\mathcal{B}} x_{e,b}^{\pi,s}(t) > 0,\ \forall e\in\mathcal{E}, \tag{8}$$

$$\sum_{e\in\mathcal{E}}\sum_{b\in\mathcal{B}} x_{e,b}^{\pi,s}(t) \le |\mathcal{B}|, \tag{9}$$

$$\frac{L_e^{\pi,s}(t)}{|\mathcal{M}|} \le \sum_{b\in\mathcal{B}} x_{e,b}^{\pi,s}(t),\ \forall e\in\mathcal{E}, \tag{10}$$

$$D(t) \le \sum_{e\in\mathcal{E}} L_e^{\pi,s}(t). \tag{11}$$

where the variables $x^{\pi} = (x_{e,b}^{\pi,s}(t) | e\in\mathcal{E}, b\in\mathcal{B}, s\in\mathcal{S})$ and $L^{\pi} = (L_{e,m}^{\pi,s}(t) | e\in\mathcal{E}, m\in\mathcal{M}, s\in\mathcal{S})$. The constraints (6)-(9) are the scheduling limits of PRB, which define that each PRB is available for one user only in the slot t and the upper bound of total allocated PRB amount is $|\mathcal{B}|$. Constraint (10) ensures that the resources preempted by punctures do not exceed the allocated resources of eMBB users. The constraint condition (11) preserves that all the arriving URLLC traffic demands are met instantly without queuing.

III. Joint Resource Scheduling Scheme

As shown in Eq. (5), the optimization of eMBB rate involves both eMBB resource allocation x^{π} and URLLC traffic placement L^{π}. The problem can be addressed by the exhaustive searching. However, its computation complexity is quite high and exponentially grows with $|\mathcal{B}|\,|\mathcal{M}|$. To decrease the computation complexity, we propose an efficient heuristic algorithm that decomposes the optimization into two procedures, i.e. eMBB resource allocation and URLLC traffic placement. With respect to eMBB resource allocation, not only the achieved eMBB rate but also the channel state and the characteristic of URLLC traffic demands are taken into account. Then utilizing the eMBB rate loss function, we opt for the the eMBB user with the least estimated loss to load URLLC traffic.

A. eMBB Resource Allocation

The resource allocation in single service scenario is generally based on channel state only. However, in scenario of multiplexing eMBB and URLLC services, the resource allocation should take the potential impact of the URLLC arrival traffic in the current slot t into account. In addition, the achieved transmission rate and peak rate of eMBB users are included in PRB allocation process as the channel state information. Therefore, we designed a weight factor denoted by:

$$\begin{aligned}\alpha_e^{\pi,s}(t) &= (1-\eta)\,R_e^{\pi}(t-1)\\ &+\eta\frac{|\mathcal{B}|\,\widehat{r_e^s}(t)}{|\mathcal{E}|}\left[1-f_e^{\pi,s}\left(\frac{|\mathcal{E}|\,L^{\pi,s}(t)}{|\mathcal{M}|\,|\mathcal{B}|}\right)\right]\end{aligned} \tag{12}$$

where $\eta\in[0,1]$, and $R_e^{\pi}(t-1)$ is the achieved transmission rate of eMBB user e until time slot $(t-1)$ under the scheduling scheme π . $\widehat{r_e^s}(t)$ is the peak rate of eMBB user e at slot t which depends on the channel state s. $f_e^{\pi,s}(\cdot)$ is the eMBB transmission rate loss function. $L^{\pi,s}(t)$ is the sum of URLLC loads in slot t under scheduling scheme π in channel state s. The specific process of eMBB resource allocation is shown as Algorithm 1. The weight factor $\alpha_e^{\pi,s}(t)$ is treated as the scheduling performance estimation of PRB b allocated to each $e\in\mathcal{E}$ based on peak rate $\widehat{r_e^s}(t)$, achieved transmission rate $R_e^{\pi}(t-1)$ of eMBB user e and URLLC loads $L^{\pi,s}(t)$. According to Algorithm 1, PRB are allocated to the eMBB users based on the proportion of the weight factor $\alpha_e^{\pi,s}(t)$. When the PRB b is assigned to user e, let $x_{e,b}^{\pi,s}(t)=1$.

Algorithm 1 eMBB Resource Allocation Algorithm

Input: D; $\mathcal{E}$; $\mathcal{B}$; $\mathcal{S}$; $\mathcal{R}=\begin{bmatrix}\hat{r}_1^1 & \cdots & \hat{r}_1^{|\mathcal{S}|}\\ \vdots & \ddots & \vdots\\ \hat{r}_{|\mathcal{E}|}^1 & \cdots & \hat{r}_{|\mathcal{E}|}^{|\mathcal{S}|}\end{bmatrix}$

Output: $x_{e,b}^{\pi,s}(t), \forall e\in\mathcal{E}, b\in\mathcal{B}$

```
Initialization x_{e,b}^{π,s}(t) = 0, ∀e ∈ E, b ∈ B
for each t ∈ T do
    for each e ∈ E do
        if t = 1 then
            α_e^{π,s}(t) = (|B| r̂_e^s(t) / |E|) (1 − f_e^{π,s}(|E| L^{π,s}(t) / (|M||B|)))
        else
            α_e^{π,s}(t) = (1 − η) R_e^π(t − 1)
                + η (|B| r̂_e^s(t) / |E|) (1 − f_e^{π,s}(|E| L^{π,s}(t) / (|M||B|)))
        end if
    end for
    Set loc = 0
    for each e ∈ E do
        Calculate N_{e,b} = (α_e^{π,s}(t) / Σ_{e'∈E} α_{e'}^{π,s}(t)) |B|
        for b = 1 to N_{e,b} do
            x_{e,b+loc}^{π,s}(t) = 1
        end for
        Update loc = loc + N_{e,b}
    end for
end for
```

B. URLLC Traffic Placement

While the PRB allocation has done according to Algorithm 1, the strategy of placing the URLLC traffic on appropriate eMBB users is the main factor affecting the eMBB rate loss. Since the amount of punctured resources is significant to eMBB transmission rate, we provided a function to describe the potential rate loss quantitatively. As the above mentioned eMBB transmission rate formula (3) is based on convex model, the potential rate loss of user $e\in\mathcal{E}$ caused by a new puncture in slot t can be formulated as:

$$\begin{aligned}loss_e^{\pi,s}(t) = \sum_{b\in\mathcal{B}} x_{e,b}^{\pi,s}(t)\,\widehat{r_e^s}(t)\Bigg[&f_e^{\pi,s}\left(\frac{L_e^{\pi,s}(t)+1}{|\mathcal{M}|\sum_{b\in\mathcal{B}} x_{e,b}^{\pi,s}(t)}\right)\\ &-f_e^{\pi,s}\left(\frac{L_e^{\pi,s}(t)}{|\mathcal{M}|\sum_{b\in\mathcal{B}} x_{e,b}^{\pi,s}(t)}\right)\Bigg]\end{aligned} \tag{13}$$

On the basis of Eq. (13), a heuristic procedure is designed to place the punctures loading URLLC traffic on the specific eMBB users in turn. As shown in Algorithm 2, if there are URLLC demands at a certain arriving rate, we will compare the potential eMBB rate loss of any $e \in \mathcal{E}$ for each new puncture. Then the eMBB user with the least potential rate loss is selected to load URLLC traffic.

Algorithm 2 URLLC Placement Algorithm

Input: D; $\mathcal{E}$; $\mathcal{B}$; $\mathcal{S}$; $\mathcal{R}$; $x_{e,b}^{\pi,s}(t), \forall e \in \mathcal{E}, b \in \mathcal{B}$
Output: $L_e^{\pi,s}(t), \forall e \in \mathcal{E}$

```
Initialization L_e^{π,s}(t) = 0, ∀e ∈ E
for each t ∈ T do
    for D(t) > Σ_{e∈E} L_e^{π,s}(t) do
        for each e ∈ E do
            Calculate loss_e^{π,s}(t)
        end for
        Select eMBB user e' ∈ E with the min loss_{e'}^{π,s}(t)
        Set L_{e'}^{π,s}(t) = L_{e'}^{π,s}(t) + 1
    end for
end for
```

The complexities of these two procedures are $\mathcal{O}(|\mathcal{E}|)$ and $\mathcal{O}(|\mathcal{E}|^{E[D(t)]})$ respectively. Considering the sporadic characteristic of URLLC traffic, the number of $E[D(t)]$ is small in reality. Hence the complexity of $\mathcal{O}(|\mathcal{E}|^{E[D(t)]})$ is acceptable to support low latency communication.

IV. Performance Evaluation

In this section, we present simulation results to assess the performance of the proposed joint resource scheduling scheme using MATLAB software. Here we illustrate the simulation parameters and significant gain of the results compared with other two baselines.

A. Assumptions and Settings in the Simulation

We suppose a single gNB wireless system serving 20 eMBB users with full buffer traffic and URLLC traffic with ftp-3 model. Specifically, the URLLC packet size is 32 bytes and the traffic arrival process follows Poisson distribution of which the arrival rate is $20 \sim 400$ $packets/ms$. There are 100 PRBs available to enable multiplexing of eMBB and URLLC traffic. The sub-carrier spacing of each PRB is 15 kHz so that the duration of slot and mini-slot are respectively 1 ms and 0.143 ms. The whole transmission duration $\mathcal{T}$ is 1 ms for 1000 slot iterations. The elements of peak rate matrix $\mathcal{R}$ are i.i.d. and uniformly distributed within $[7, 13] Mbps$, and the number of channel states is 100.

B. Performance Results

The performance of the proposed joint resource scheduling scheme is evaluated through the sum rate of all eMBB users and corresponding empirical cumulative distribution function (ECDF). We consider two comparative scheduling scheme:

- **Resource Proportional (RP) Scheme**: RP adopts the eMBB resource allocation strategy based on the achieved eMBB rate till slot $t-1$ and the peak rate in slot t. Its puncturing placement for loading URLLC traffic is proportional to the allocated resources of eMBB users.
- **Estimated Loss Puncturing (ELP) Scheme**: ELP utilizes the same eMBB resource allocation strategy as adopted by RP. Taking the eMBB expected rate loss due to URLLC traffic into account, ELP punctures eMBB users with least estimated loss.

The allocated PRBs and the placed URLLC traffic of each eMBB user is illustrated in Fig. 2. As a result of taking the URLLC traffic arrival characteristic into account, the proposed scheme finds an eMBB resource allocation different from that of the ELP and RP schemes in the same channel states. Fig. 3 gives the ECDF of eMBB data rate at the URLLC arrival rate of $20 \sim 400$ $packets/ms$. It is shown that the proposed scheme presents the best performance with the help of the joint scheduling of eMBB resource allocation and URLLC traffic placement considering the channel state and the characteristic of the URLLC traffic demands. The ELP scheme offers the second best performance due to its URLLC traffic placement strategy with the estimated eMBB rate loss. The RP scheme performs worst due to the resource proportional puncturing placement and the eMBB resource allocation which neglects the URLLC traffic arrival characteristic. As shown in Fig.

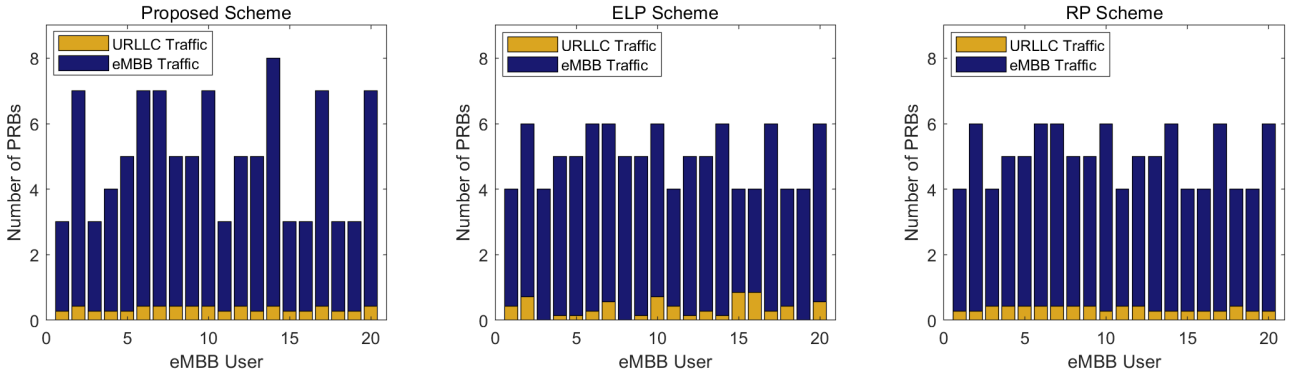

Fig. 2. Allocated PRBs of eMBB users and distributed URLLC traffic.

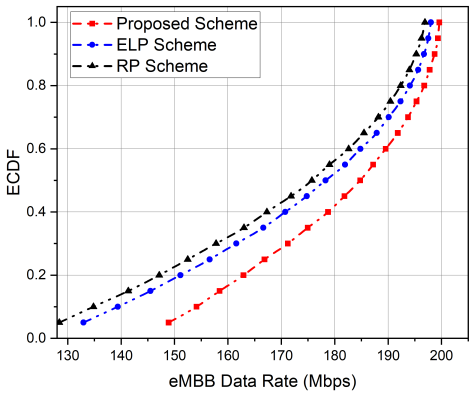

Fig. 3. Comparison of ECDF for achievable rate of eMBB users.

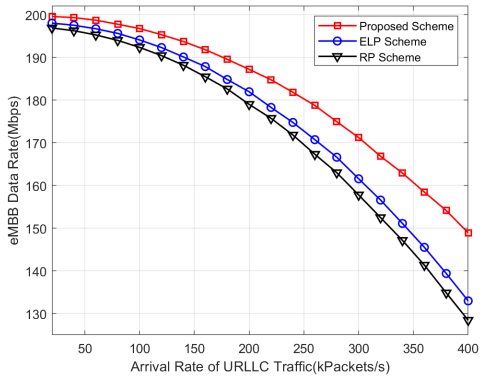

Fig. 4. Comparison of eMBB rate for different URLLC arrival traffic.

4, the overall trend of the eMBB rate performance declines with increased URLLC traffic arrival rate as a result of more time-frequency resources preempted by URLLC traffic. The proposed method provided up to 11.9% ∼15.8% eMBB rate gain at the URLLC traffic arrival rate of 400 $packets/ms$ compared with ELP and RP. The results demonstrate that the proposed scheme enable an improved eMBB rate via the joint resource scheduling of eMBB and URLLC traffic. When the proposed scheduling scheme is applied in reality, the accuracy of channel state information does have some effect on the eMBB performance compared to the above simulations, but the general trend of the three schemes is similar to the simulation results.

V. Conclusion

In this paper, we present an effective solution for multiplexing of URLLC and eMBB traffic in 5G wireless networks. We have formulated an optimization problem to improve performance of eMBB rate impacted by sporadic URLLC arrival traffic. Specifically, a practical convex model is utilized to describe the relationship between the potential eMBB transmission rate loss and the preempted resources for puncturing. Further, a joint scheduling of eMBB and URLLC traffic is proposed, where both the URLLC traffic arrival characteristics and the channel state are considered in eMBB resource allocation. The simulation results demonstrate that the proposed scheme achieves better performance in terms of eMBB transmission rate than two baselines.

Acknowledgment

This work is supported by the National Natural Science Foundation of China under Grant No. 61971069, 61801051, Beijing Natural Science Foundation under Grant No. L202003, and the Key RD Program Projects in Shanxi Province under Grant No. 2019ZDLGY07-10.

References

[1] M. Alsenwi, N. H. Tran, M. Bennis, A. Kumar Bairagi, and C. S. Hong. embb-urllc resource slicing: A risk-sensitive approach.IEEE Communications Letters, 23(4):740–743, April 2019.

[2] R. Abreu, T. Jacobsen, K. Pedersen, G. Berardinelli, and P. Mogensen, "System level analysis of embb and grant-free urllc multiplexing in uplink," in 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring). IEEE, 2019, pp. 1–5.

[3] M. Bennis, M. Debbah, and H. V. Poor. Ultrareliable and low-latency wireless communication: Tail, risk, and scale.Proceedings of the IEEE, 106(10):1834–1853, Oct 2018.

[4] K. S. Kim, D. K. Kim, C. Chae, S. Choi, Y. Ko, J. Kim, Y. Lim, M. Yang, S. Kim, B. Lim, K. Lee, and K. L. Ryu. Ultrareliable and low-latency communication techniques for tactile internet services. Proceedings of the IEEE, 107(2):376–393, Feb 2019.

[5] P. Popovski, C. Stefanović, J. J. Nielsen, E. de Carvalho, M. Angjelichinoski, K. F. Trillingsgaard, and A.-S. Bana, "Wireless access in ultra-reliable low-latency communication (URLLC),"IEEE Trans. Commun., vol. 67, no. 8, pp. 5783–5801, Aug. 2019.

[6] G. Pocovi, K. I. Pedersen, and P. Mogensen. Multiplexing of latency-critical communication and mobile broadband on a shared channel. In2018 IEEE Wireless Communications and Networking Conference (WCNC), pages 1–6, April 2018.

[7] A. Anand, G. de Veciana, and S. Shakkottai, "Joint scheduling of URLLC and eMBB traffic in 5G wireless networks," inProc. INFOCOM, May 2018, pp. 1970–1978.

[8] A. Gosh, "5G new radio (NR): Physical layer overview and performance," inProc. IEEE Commun. Theory Workshop, May 2018, pp. 1–38.

[9] M. Alsenwi and C. S. Hong, "Resource scheduling of URLLC/eMBB traffics in 5G new radio: A punctured scheduling approach," inProc. Korean Comput. Congr. (KCC), Jun. 2018, pp. 1271–1273.

[10] R. Abreuet al., "On the multiplexing of broadband traffic and grant-free ultra-reliable communication in uplink," inProc. IEEE 89th Veh. Technol. Conf. (VTC Spring), Jan. 2019, pp. 1–6.

[11] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Ultra-reliable and low-latency communications in 5G downlink: Physical layer aspects,"IEEE Wireless Commun., vol. 25, no. 3, pp. 124–130, Jun. 2018.

[12] J. J. Nielsen, R. Liu, and P. Popovski, "Ultra-reliable low latency communication using interface diversity,"IEEE Trans. Commun., vol. 66, no. 3, pp. 1322–1334, Mar. 2018.

[13] Z. Zhou, P. Liu, J. Feng, Y. Zhang, S. Mumtaz, and J. Rodriguez, "Computation resource allocation and task assignment optimization invehicular fog computing: A contract-matching approach,"IEEE Trans. Veh. Technol., vol. 68, no. 4, pp. 3113–3125, Apr. 2019.

[14] A. Manzoor, S. M. A. Kazmi, S. R. Pandey and C. S. Hong, "Contract-Based Scheduling of URLLC Packets in Incumbent EMBB Traffic," in IEEE Access, vol. 8, pp. 167516-167526, 2020.

[15] S. M. A. Kazmi, N. H. Tran, T. M. Ho, A. Manzoor, D. Niyato, and C. S. Hong, "Coordinated Device-to-Device communication with non-orthogonal multiple access in future wireless cellular networks,"IEEE Access, vol. 6, pp. 39860–39875, Jun. 2018.

[16] T. M. Ho, N. H. Tran, S. M. A. Kazmi, Z. Han, and C. S. Hong, "Wireless network virtualization with non-orthogonal multiple access," inProc. IEEE/IFIP Netw. Oper. Manage. Symp. (NOMS), Apr. 2018, pp. 1–9.

[17] J. Li and X. Zhang, "Deep Reinforcement Learning-Based Joint Scheduling of eMBB and URLLC in 5G Networks," in IEEE Wireless Communications Letters, vol. 9, no. 9, pp. 1543-1546, Sept. 2020.
[18] Q. Shang, F. Liu, C. Feng, R. Zhang and S. Zhao, "A BP Neural Network Based Punctured Scheduling Scheme Within Mini-slots for Joint URLLC and eMBB Traffic," 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2019, pp. 1-5.
[19] L. Zhang, J. Tan, Y. Liang, G. Feng and D. Niyato, "Deep Reinforcement Learning-Based Modulation and Coding Scheme Selection in Cognitive Heterogeneous Networks," in IEEE Transactions on Wireless Communications, vol. 18, no. 6, pp. 3281-3294, June 2019.
[20] A. K. Bairagi et al., "Coexistence Mechanism Between eMBB and uRLLC in 5G Wireless Networks," in IEEE Transactions on Communications, vol. 69, no. 3, pp. 1736-1749, March 2021.
[21] A. A. Esswie and K. I. Pedersen, "Opportunistic Spatial Preemptive Scheduling for URLLC and eMBB Coexistence in Multi-User 5G Networks," in IEEE Access, vol. 6, pp. 38451-38463, 2018.
[22] A. Pradhan and S. Das, "Joint Preference Metric for Efficient Resource Allocation in Co-Existence of eMBB and URLLC," 2020 International Conference on COMmunication Systems NETworkS (COMSNETS), 2020, pp. 897-899.
[23] A. Anand, G. de Veciana and S. Shakkottai, "Joint Scheduling of URLLC and eMBB Traffic in 5G Wireless Networks," in IEEE/ACM Transactions on Networking, vol. 28, no. 2, pp. 477-490, April 2020.