Reinforcement Learning-Based Joint Power and Resource Allocation for URLLC in 5G

基于增强学习的5G URLLC联合能量和资源分配

> *Elsayed M, Erol-Kantarci M. Reinforcement Learning-Based Joint Power and Resource Allocation for URLLC in 5G[C]//2019 IEEE Global Communications Conference (GLOBECOM). IEEE, 2019: 1-6.*

一分钟概括

问题

针对5G新无线电(NR)共享信道上，超可靠低延迟通信(URLLC)用户和增强型移动宽带(EMBB)用户的复用问题

解决方案

提出了一种基于Q学习的联合功率和资源分配算法。在不影响eMBB用户吞吐量的前提下，提高了URLLC用户的体验。 减轻小区间干扰以及改善传输和调度延迟。

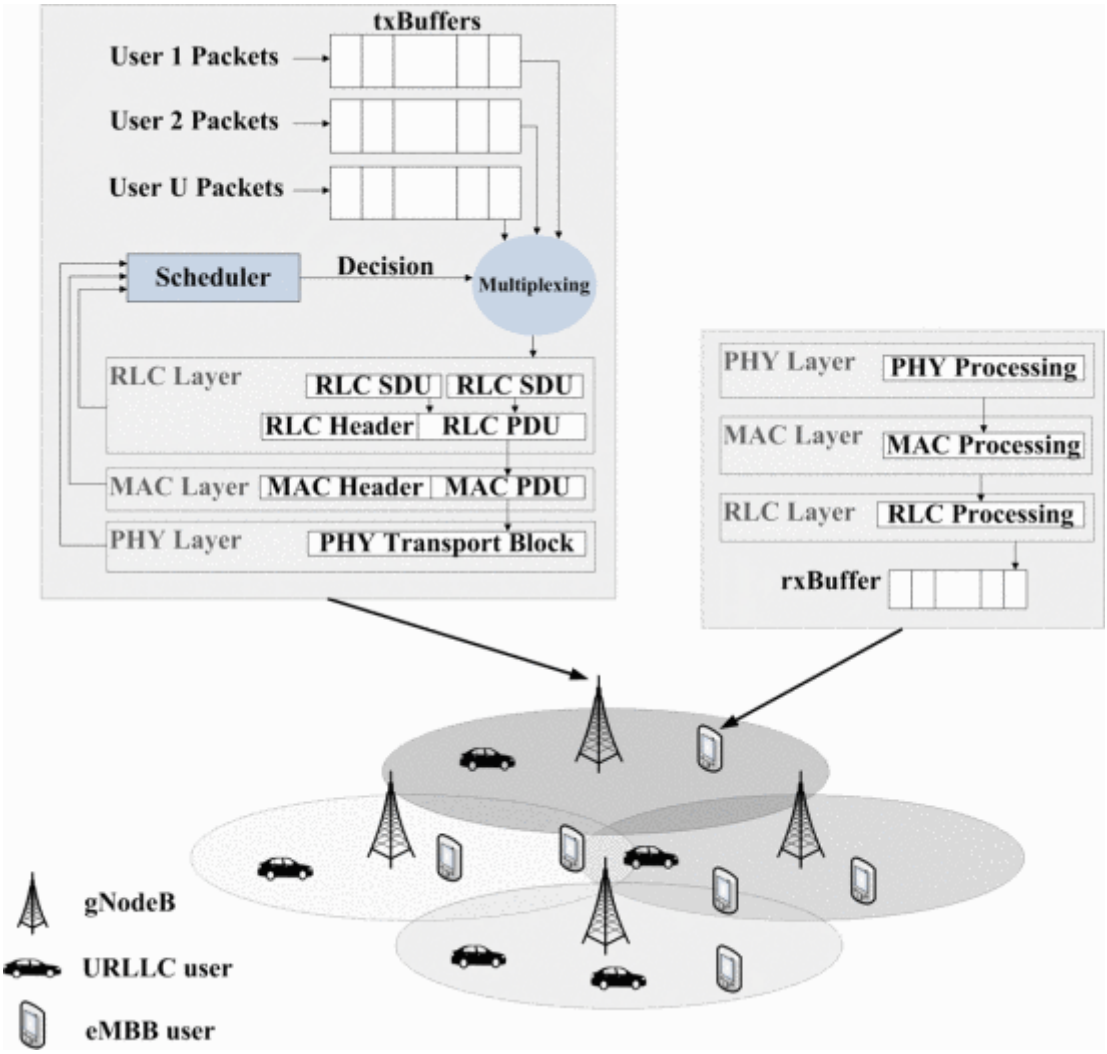
仿真

将本算法与Baseline算法进行比较, 指标有**延迟、PDR、吞吐量**

结论

提出了一种基于Q学习的联合功率和资源分配技术，以改善5G中URLLC用户的体验。 该算法在URLLC用户的可靠性和延迟方面都优于基线算法，同时为eMBB用户实现了更高的吞吐量。

系统模型



gNodeB：基站

URLLC user：低延迟高可靠性通信需求的用户

eMBB user：增强型移动宽带用户

每个用户相对应的流量在gNodeB维护的传输缓冲区中排队。gNodeB会触发每个下行链路调度器的下行调度程序，以执行传输缓冲区中待处理流量的资源块分配。

延迟和可靠性的定义

$$D = \tau + D_{tx} + \sum_{r=1}^{n} D_{r,harq}, \tag{1}$$

其中τ是排队延迟，D_{tx}是传输延迟，而$D_{r,harq}$是HARQ重传的往返延迟，其中n是重传次数，为了保证较低的延迟，在本文中设置为1。

第j个gNodeB上的第i个用户的传输延迟可以表示为：

$$D_{tx,i,j} = \frac{S_{i,j}}{\sum_{k=1}^{K} \omega_k log_2 \left(1 + \frac{p_{k,j} d_{k,i,j} h_{k,i,j}}{\omega_k N_0 + \sum_{\substack{m \in \mathcal{J} \\ m \neq j}} p_{k,m} d_{k,i,m} h_{k,i,m}}\right)}, \tag{2}$$

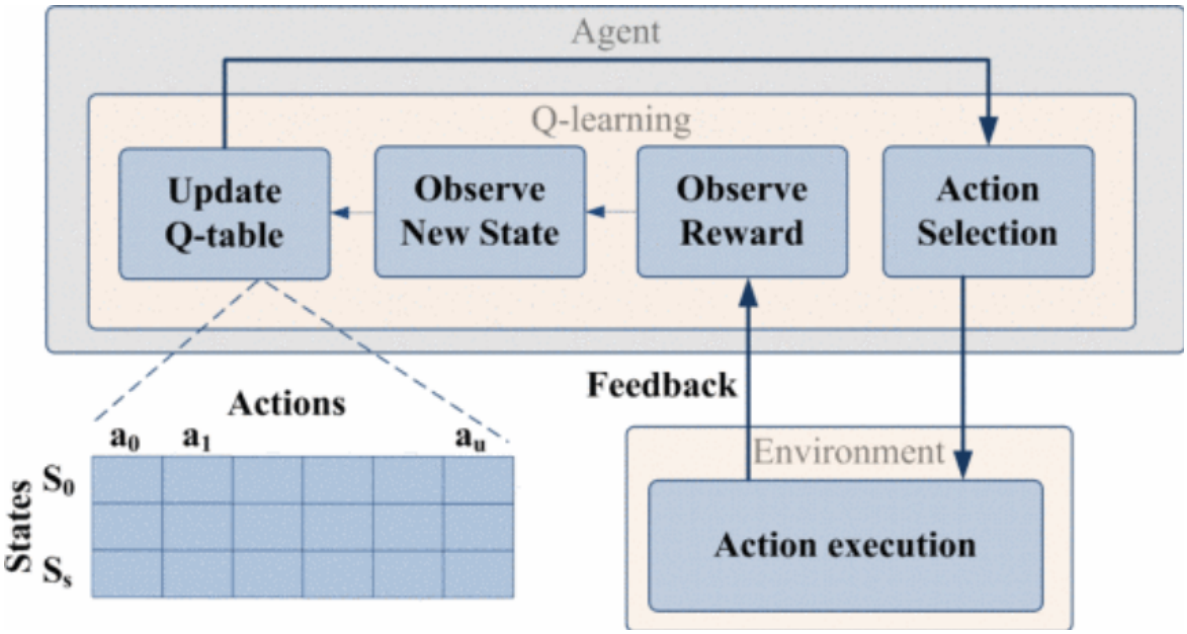
其中Si,j是链路(i,j)的分组大小，ω_k是第k个RBG的带宽，N_0是加性高斯噪声单侧功率谱密度。 pk,j是第k个gNodeB在第k个RBG上的发射功率，$h_{k,i,j}$是信道系数，$d_{k,i,j}$是RBG的链路分配指示符(k，i，j)。$p_{k,m}$是第m个干扰gNodeB的发射功率，$h_{k,i,m}$是信道系数，$d_{k,i,m}$是链路的分配指示符(k，i，m)。

为了平衡URLLC用户的延迟和可靠性，需要进行SINR估计、功率分配和RBG分配。在接下来的章节中，我们提出了一种基于Q-learning的**功率和RBG分配**联合优化算法，用于URLLC的延迟和可靠性改进。

基于Q学习的URLLC低延时高可靠性算法(LLHRQ)

gNodeB都采用多主体强化学习，就是Q学习算法

图2给出了Q-learning框架运行的一般框图。



> *Q-learning 由元组定义:agent, state, actions, reward, policy, and Q-value。*
> *Q-Learning是强化学习算法中value-based的算法，Q即为Q（s,a）就是在某一时刻的s 状态下(s∈S)，采取 动作a (a∈A)动作能够获得收益的期望，环境会根据agent的动作反馈相应的回报reward r，所以算法的主要思想**就是将State与Action构建成一张Q-table来存储Q值，然后根据Q值来选取能够获得最大的收益的动作。***
>
> *https://blog.csdn.net/qq_30615903/article/details/80739243*

因此，Q-learning使用如下迭代更新来估计已访问状态-动作对的Q值:

$$Q(s^{(t)},a^{(t)}) \leftarrow Q(s^{(t)},a^{(t)}) + \alpha[r^{(t)} + \gamma \max_a Q^{old}(s^{(t+1)},a) - Q(s^{(t)},a^{(t)})], \quad (3)$$

该Q学习算法，以提高可靠性并最小化URLLC用户的延迟。

1) Agents

gNodeBs.

2) Actions

- 每个gNodeB对其连接的用户采取的联合功率和资源块分配。
- 第j个gNodeB对第k个RBG的动作可以定义为A_k，i={D_k，i，P_k，j}，其中$D_{k,i}$={$D_{k,i,j}$：i∈U}是每个用户的RBG的分配指示符的向量，即，如果将第k个RBG分配给第i个用户，则$D_{k,i,j}$=1，否则为0，并且$P_{k,j}$是分配的功率

3) States

- 在没有Agent协作的情况下，来自环境的反馈应该起到以状态为代表的作用。 特别地，我们通过估计每个用户的SINR值来观察干扰对URLLC用户的影响。 这量化了URLLC用户的可靠性。 因此，两个状态设计如下：

$$S_{k,j}^{(t)} = \begin{cases} S_0 & \lfloor \frac{1}{R} \sum_{i_R \in \mathcal{R}} \gamma_{k,i_R,j}^{(t-1)} \rfloor \geq \gamma_{th}, \\ S_1 & Otherwise. \end{cases} \tag{4}$$

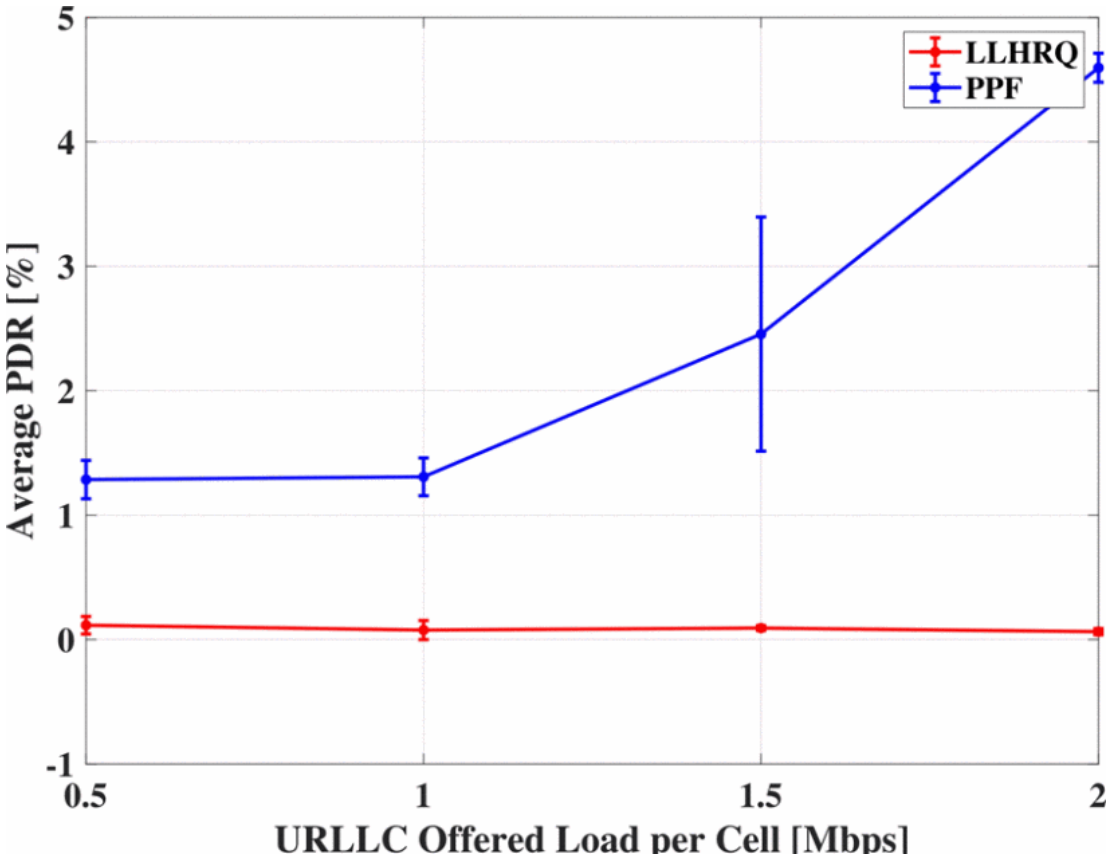
4) Reward

奖励函数：

$$R_{k,j}^{(t)} = \begin{cases} 1 - \max_{i \in \mathcal{R}} (\tau_{i_R,j}^{(t-1)})^2 & \frac{1}{R} \sum_{i_R \in \mathcal{R}} \gamma_{k,i_R,j}^{(t-1)} \geq \gamma_{th}, \\ -1 & otherwise, \end{cases} \tag{5}$$

性能评估

图4评估了可靠性和延迟之间的关系



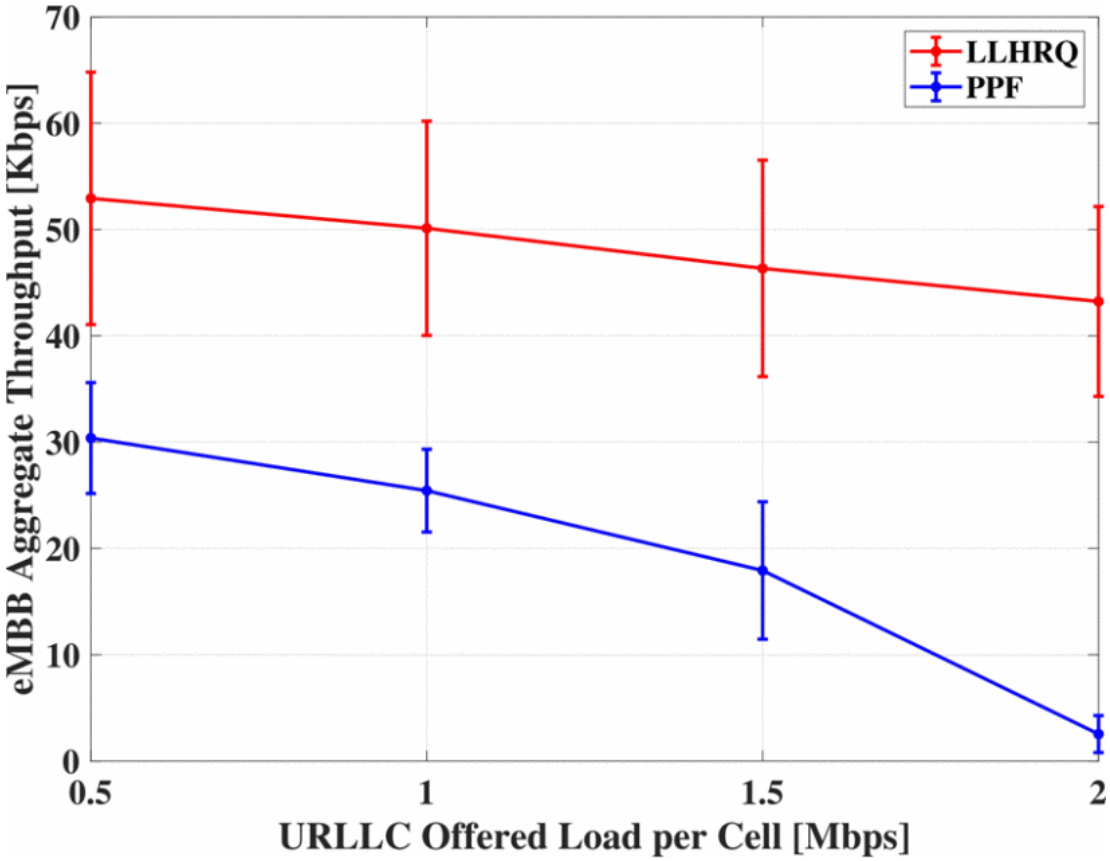


图5展示了在URLLC用户的各种流量负载下eMBB用户的吞吐量。

结论

提出了一种基于Q学习的联合功率和资源分配技术，以改善5G中URLLC用户的体验。 该算法在URLLC用户的可靠性和延迟方面都优于基线算法，同时为eMBB用户实现了更高的吞吐量。