Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Uplink puncturing for mixed URLLC and eMBB services in 5G-based IWNs: a model-aided DRL method*

Jingfang DING[†1], Meng ZHENG[†‡1], Haibin YU[†‡2,3], Yitian WANG[2,3,4], Chi XU[2,3]

[1]*School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China*

[2]*State Key Laboratory of Robotics, Chinese Academy of Sciences, Shenyang 110016, China*

[3]*Key Laboratory of Networked Control Systems, Chinese Academy of Sciences, Shenyang 110016, China*

[4]*University of Chinese Academy of Sciences, Beijing 100049, China*

[†]E-mail: dingjingfang@cse.neu.edu.cn; zhengmeng@cse.neu.edu.cn; yhb@sia.cn

Received Mar. 18, 2025; Revision accepted July 22, 2025; Crosschecked

Unedited

Abstract: The coexistence of Ultra-Reliable Low-Latency Communication (URLLC) and Enhanced Mobile Broadband (eMBB) services in 5th Generation (5G)-based Industrial Wireless Networks (IWNs) poses significant resource slicing challenges due to their inherent performance requirement conflicts. To address this challenge, this paper proposes a puncturing method that utilizes a model-aided Deep Reinforcement Learning (DRL) algorithm for URLLC over eMBB services in uplink 5G networks. First, a puncturing-based optimization problem is formulated to maximize the accumulated eMBB rate under strict URLLC latency and reliability constraints. Next, we design a Random Repetition Coding-based Contention (RRCC) scheme for sporadic URLLC traffic and derive its analytical reliability model. To jointly optimize the scheduling parameters of URLLC and eMBB, a DRL solution based on the reliability model is developed, which is capable of dynamically adapting to changing environments. The accelerated convergence of the model-aided DRL algorithm is demonstrated using simulations, and the superiority in resource efficiency of the proposed method over existing approaches is validated.

Key words: Uplink 5G networks; eMBB; URLLC; Resource slicing; Puncturing; Deep reinforcement learning
https://doi.org/10.1631/FITEE.2500173 **CLC number:**

1 Introduction

Industrial Wireless Networks (IWNs) serve as critical enablers of the seamless integration of the Internet of Things in the Industry 4.0 ecosystem and facilitate interconnections among varied devices—such as sensors (or cameras), actuators, Automated Guided Vehicles (AGVs), Robotic Manipulators (RMs), and virtual reality interaction systems—to ensure the efficient and uninterrupted operation of smart manufacturing processes (Sardar et al., 2024). However, the diverse communication requirements of these devices have led to the emergence of heterogeneous services in industrial automation scenarios, primarily Ultra-Reliable Low-Latency Communication (URLLC) and Enhanced Mobile Broadband (eMBB), both of which are supported by the 5th Generation (5G) New Radio (NR) (Ji et al., 2018). Specifically, URLLC services, designed for applications requiring low latency and high reliability, must achieve a 99.999% reliability when transmitting 32 bytes of data within a user-plane latency of 1 ms, as defined by the 3rd Generation Partnership Project (3GPP) (Ma et al., 2019). In contrast, eMBB services are tailored for high-bandwidth applications and aim to deliver ultra-high data rates with a

‡ Corresponding authors

* Project supported by the Liaoning Revitalization Talents Program (Nos. XLYC2203148 and XLYC2403062) and the National Natural Science Foundation of China (Nos. 92267108 and 62173322)

ORCID: Jingfang DING, https://orcid.org/0000-0002-0019-8200; Meng ZHENG, https://orcid.org/0000-0003-1674-8577; Haibin YU, https://orcid.org/0000-0002-1663-2956

© Zhejiang University Press 2025

moderate reliability level of 99.9% (Popovski et al., 2018). Such service diversity introduces significant complexities in resource slicing and traffic scheduling, requiring advanced solutions to optimize network performance.

Unlike downlink transmission, where the Base Station (BS) has full control over data packets, uplink transmission presents a unique challenge because the BS lacks precise information about whether User Equipment (UE) has packets ready to transmit. Moreover, uplink access methods exhibit significant differences across various traffic types. In 5G, three access methods are principally adopted: Grant-Based (GB) access, Semi-Persistent Scheduling (SPS), and Grant-Free (GF) access (Nomeir et al., 2023). GB access is suitable for general eMBB traffic without strict latency requirements because it relies on dynamic scheduling requests. SPS, widely used for fully buffered eMBB traffic and periodic URLLC traffic, reduces signaling overhead by pre-allocating resources within specific time intervals. GF access operates in an "arrive-and-go" manner with minimal signaling overhead, making it particularly efficient for sporadic URLLC traffic. This paper focuses mainly on the uplink transmissions of fully buffered eMBB traffic and sporadic URLLC traffic. Unlike schedulable periodic traffic, sporadic traffic, with its uncertain characteristics, requires more in-depth research.

To date, enabling the coexistence of eMBB and URLLC services in the uplink has primarily depended on the following methods:

- **Puncturing methods,** which allow URLLC transmissions to preempt the time-frequency resources of ongoing eMBB transmissions, include two types: (a) informed puncturing, where the BS receives advance notification of URLLC packets and proactively cancels scheduled eMBB transmissions to accommodate URLLC traffic, and (b) direct puncturing, where URLLC transmissions are directly overlaid on eMBB transmissions through Power Control (PC) mechanisms (3GPP, 2018). Waiting for notifications may not be suitable for packets with stringent latency requirements. Hence, direct puncturing is more appropriate for real-time URLLC traffic.

- **Reservation methods,** which pre-allocate dedicated time-frequency resources for URLLC and eMBB traffic, also employ two types (Alsenwi et al., 2021): (a) semi-static reservation, where the BS periodically broadcasts frame structure configurations, and (b) dynamic reservation, which frequently updates frame structure information through UE control channels. For sporadic URLLC traffic with a low generation rate, reservation methods may lead to substantial resource wastage, while the dynamic reservation incurs higher control overhead.

- **Non-Orthogonal Multiple Access (NOMA)-based methods,** which enable multiple UEs to share communication resources in time-frequency domains (Liu et al., 2024), with decoding at the receiver via power or code domain multiplexing. Nonetheless, NOMA imposes stringent receiver requirements, necessitating advanced Successive Interference Cancellation techniques, which pose significant implementation challenges in industrial automation scenarios.

Inspired by the previous discussion, we propose a puncturing method based on a model-aided Deep Reinforcement Learning (DRL) algorithm to address the coexistence problem of URLLC and eMBB services in uplink 5G networks. The main contributions of this work are summarized as follows:

- We study a puncturing method that allows URLLC transmissions to puncture ongoing eMBB transmissions across diverse resource unit (RU) scales and then formulate the puncturing problem as a constrained optimization problem, aiming to maximize the accumulated transmission rate of eMBB UEs while strictly satisfying the latency and reliability requirements of URLLC UEs.

- We designed a Random Repetition Coding-based Contention (RRCC) scheme for sporadic URLLC traffic that enables randomly generated URLLC packets to immediately select any K available RUs for repetitive transmissions before their deadlines. An analytical reliability model is derived for the RRCC scheme, demonstrating that the proposed scheme requires fewer

Unedited

repetitions compared to existing schemes while achieving the same reliability level.

- We develop a DRL solution based on the aforementioned reliability model, which jointly optimizes the repetition number of URLLC traffic and the scheduling list of eMBB traffic, achieving effective responsiveness to dynamic environments.

- Using simulations, we demonstrate the improved convergence speed of the model-aided DRL algorithm and showcase the resource efficiency advantages of the proposed method in comparison with existing reservation methods.

2 Related works

2.1 GF access for URLLC

Among GF access methods in 5G, K-repetition, which involves transmitting K repetitions of a packet without awaiting feedback, has been widely considered as a promising solution in uplink transmission, effectively enhancing reliability for URLLC services with strict deadlines. For the first time, Singh et al. (2018) conducted collision probability analysis of K-repetition. Subsequently, a succession of representative variants of K-repetition has come into being. For instance, Lee et al. (2018) designed various repetition patterns for UEs to reduce collision probability. Yuan et al. (2020) combined SPS for initial transmission with GF for retransmissions, effectively reducing signaling overhead. Furthermore, Song et al. (2022) explored an Optimized Repetition Scheme (ORS), where each UE transmits K repetitions orthogonally in the frequency domain with Orthogonal Frequency-Division Multiple Access (OFDMA) technology. More recently, Zhao et al. (2023) introduced a Semi-Probabilistic Repetition (SPR) scheme, leveraging Multi-User Detection (MUD) to enhance transmission reliability. Ding et al. (2024) proposed the q-Consecutive Repetition Coding-based Contention (q-CRCC) scheme incorporating MUD. However, both SPR and q-CRCC are limited to transmitting repetitions in consecutive Transmission Time Intervals (TTIs). Moreover, both Ding and Zheng (2022) and Elayoubi et al. (2019) focused on the real-time and reliable transmission of sporadic URLLC packets within periodic time intervals, where Ding and Zheng (2022) considered the resource slicing method to allocate an exclusive part of resources for single-antenna UEs with URLLC packets, and Elayoubi et al. (2019) proposed a flexible retransmission scheme for URLLC packets.

Although rich contributions concerning GF access for URLLC exist, the works above may not be adequate to deal with massive collisions when the network scales. As a result, there remains an urgent demand for a scalable K-repetition coding scheme that can reduce collisions efficiently.

2.2 Coexistence of uplink URLLC and eMBB

2.2.1 Model-based coexistence methods

Model-based coexistence methods for uplink URLLC and eMBB can be categorized into three types: puncturing methods, reservation methods, and NOMA-based methods.

In the context of puncturing methods, informed puncturing has been explored in several studies. Zaki-Hindi et al. (2020) proposed a preemptive method based on the listen-before-talk mechanism, where URLLC packets are transmitted with high power when their delays are approaching the deadlines. Xiao et al. (2020) and Han et al. (2022) investigated uplink cancellation indication mechanisms utilizing PC, relying on GB scheduling. Concerning direct puncturing, Abreu et al. (2019) proposed an overlay method for URLLC over eMBB based on PC trade-offs and evaluated its performance through extensive system-level simulations. Khodakhah et al. (2023) also considered a puncturing method for URLLC over eMBB and assessed URLLC performance using the age of information.

Regarding reservation methods, Kassab et al. (2018) proposed an orthogonal multiple access scheme in which a mini-slot is reserved for URLLC at the end of each scheduling period. Nomeir et al. (2021) introduced an uplink scheduling method that broadcasts available resources for mixed GB eMBB and GF URLLC at the beginning of each TTI, where URLLC can transmit multiple sets of K repetitions based on feedback. The aforementioned approach falls under semi-static reservation. In contrast, Feng et al. (2020) implemented a dynamic resource allocation strategy based on the queue backlog status.

Unedited

In recent years, there has been growing research on NOMA-based methods (Liu et al., 2024; Khodakhah et al., 2023; Kassab et al., 2018), which enhance spectrum utilization and resolve resource conflicts between heterogeneous services. However, such technology has not been extensively explored in industrial automation scenarios. Direct puncturing, with its minimal resource wastage and signaling overhead, is of greatest interest to us.

2.2.2 DRL-based coexistence methods

DRL algorithms have demonstrated significant advantages in solving resource slicing problems. Through interaction with the environment, DRL can autonomously learn optimal resource slicing strategies, making it well-suited for complex and dynamically changing environments. Nomeir et al. (2023) proposed a joint optimization method for uplink eMBB and URLLC traffic based on DRL and Deep Neural Networks (DNNs). First, this method reserves resources for URLLC within each TTI to meet performance requirements and then optimally allocates the remaining resources to eMBB UEs. Jiang et al. (2023) studied a DRL-based resource slicing approach that incorporates preemptive puncturing for uplink transmission. However, the harvest-then-transmit protocol adopted in this approach penalizes latency and cannot account for the reliability constraints of URLLC services.

Nevertheless, research on the coexistence of heterogeneous services in uplink transmission using machine learning is still rather limited. In contrast, there is a wealth of learning-based studies concerning downlink transmission research. For instance, Huang et al. (2020), Li and Zhang (2020), and Alsenwi et al. (2021) explored preemptive (puncturing) scheduling based on DRL, while Tian et al. (2024) and Filali et al. (2022) investigated reservation scheduling using DRL. Setayesh et al. (2022) used hierarchical deep learning to combine reservation scheduling with dedicated bandwidth and preemptive scheduling with shared bandwidth. These studies highlight the potential of DRL in addressing more complex uplink resource slicing problems. Meanwhile, Alsenwi et al. (2021) proposed an optimization-aided DRL framework. By combining model-free and model-based methods, this framework adapts effectively to dynamic environments and alleviates the slow convergence of DRL, offering a promising direction for future research.

3 System model

3.1 An industrial automation scenario with 5G-based IWNs

We consider a typical automated factory floor with 5G network coverage, as illustrated in Fig. 1, where a 5G BS, an edge server, AGVs, RMs, and cameras are equipped. The AGVs handle material transportation tasks, while the RMs perform assembly tasks. Both AGVs and RMs periodically or sporadically establish URLLC connections with the BS to upload sensor data and receive control commands. The cameras are used for monitoring the factory environment and continuously transmit high-definition video to the BS. The edge server, which manages network resource allocation and centralized control of the entire factory floor, is deployed at the BS side and connected to the BS via optical fibers (Zhou et al., 2025).

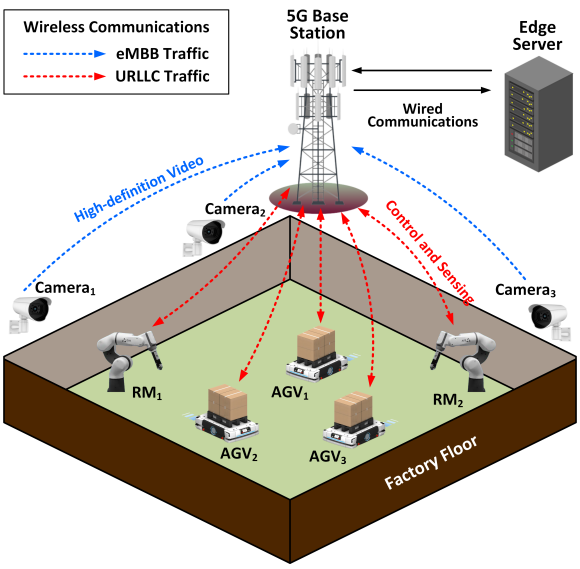

Fig. 1 Example of an industrial automation scenario with 5G-based IWNs: a fully automated factory floor. 5G, 5th Generation; IWNs, Industrial Wireless Networks.

This paper focuses on the uplink 5G network and primarily considers two types of traffic in this scenario: (a) sporadic URLLC traffic from AGVs and RMs to the BS and (b) fully buffered eMBB traffic from cameras to the BS. More specifically, we assume that the system consists of N_u URLLC UEs

Unedited

with a packet size of b_u and N_e eMBB UEs with a packet size of b_e. The summary of notations used in this work is presented in Table 1.

Table 1 Summary of notations

Notation	Definition
W	Available bandwidth
μ	Numerology parameter
τ	Duration of timeslot
f_b	Subcarrier spacing
w	Bandwidth of an RB
η	Spectral efficiency of the MCS
N_u (N_e)	Number of URLLC (eMBB) UEs
b_u (b_e)	Size of URLLC (eMBB) packets
τ_S (τ_L)	Duration of S-TTI (L-TTI)
R_S (R_L)	Number of RBs in an S-RU (L-RU)
B_S (B_L)	Bandwidth of an S-RU (L-RU)
M_S (M_L)	Number of S-RUs (L-RUs) within W
P_a	Activity probability of URLLC UEs
λ	Generation rate of URLLC packets
D	Deadline of URLLC packets
T_u	TTI numbers within the deadline
R_u	Average reliability of URLLC UEs
δ	Lower limit of R_u
p	Packet error rate
K	Repetition number
T	Duration of scheduling period
T_e	TTI numbers within the period
P_e	Transmit power of eMBB UEs
σ^2	Noise power
R_e	Average transmission rate of eMBB UEs
ξ	Rician-distributed random variable
β	Rician factor
$S^j_{m,n}$	Scheduling list
$r_{m,n}$	Puncturing list
SNR^j	SNR of the jth eMBB UE
h^j	Channel gain of the jth eMBB UE
c_1	Penalty factor
c_2	Scaling factor
$r_{\min}$	Lower bound of target reward
$r_{\max}$	Upper bound of target reward

- For URLLC UEs, the sporadic traffic follows a mutually independent Poisson distribution which is widely used for random packet generation or arrival processes (Singh et al., 2018; Abreu et al., 2019; Zhao et al., 2023; Ding et al., 2024). Therefore, the probability P_a of event X that a URLLC UE experiences random packet generation within a TTI is given by $P_a(\lambda, 1) = P(X > 0) = 1 - e^{-\lambda}$. Here, λ represents the generation rate, defined as the average number of random generation events within a TTI. The deadline for URLLC packets is set to D ms. We define the transmission reliability metric as R_u (the detailed calculation is provided in Section 4), which represents the average success probability of URLLC packets, where success means no collisions and no channel errors. The reliability requirements for URLLC transmissions within the deadline are specified as δ, necessitating $R_u \geq \delta$.
- For eMBB UEs, the traffic follows a fully buffered model, ensuring continuous packet availability for each eMBB UE to be scheduled across a certain number of TTIs. We employ T ms as the resource scheduling period for eMBB transmissions. The performance of eMBB transmissions is typically evaluated using the average transmission rate R_e (the detailed calculation is provided in Section 5), which is particularly relevant for bandwidth-intensive applications.

Unedited

3.2 Wireless channel impairments

The presence of multipath propagation in wireless channels leads to fading phenomena, which substantially affect transmission rates. Such fading is commonly modeled by either Rician or Rayleigh distributions (Alsenwi et al., 2021). We represent the Rician-distributed random variable as $\xi_\beta(t)$. The Rician fading channel model incorporates a dominant Line-of-Sight (LOS) component along with multiple reflected Non-Line-of-Sight (NLOS) components. The NLOS components are characterized by a Rayleigh distribution and can be represented as a complex Gaussian random variable $\frac{(X_1(t)+jX_2(t))}{\sqrt{2}}$, where X_1 and X_2 are independent and identically distributed (i.i.d.) Gaussian random variables with $\mathcal{N}(0,1)$. The amplitude of the NLOS component is $\frac{|X_1(t)+jX_2(t)|}{\sqrt{2}}$. β denotes the Rician factor, and thus,

$$\xi_\beta(t) = \sqrt{\frac{\beta}{\beta+1}} + \sqrt{\frac{1}{\beta+1}}\frac{|X_1(t)+jX_2(t)|}{\sqrt{2}}. \quad (1)$$

As $\beta \to \infty$, the distribution converges to Gaussian, representing only the LOS component; as $\beta \to 0$, the distribution approaches Rayleigh, representing only the NLOS component.

Due to channel instability, wireless communications are particularly susceptible to packet loss, consequently compromising transmission reliability. We define the packet error rate as p, which can be empirically determined through long-term channel

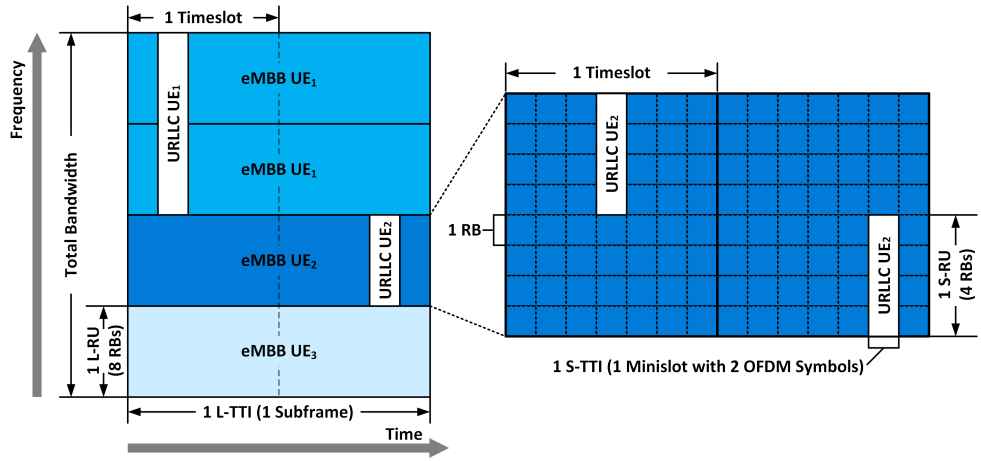

Fig. 2 Example of 5G resource configuration for heterogeneous services. 5G, 5th Generation.

measurements. For simplicity, this work adopts a fixed p for each URLLC packet throughout its deadline duration of D ms. This quasi-static assumption holds when D is shorter than the channel coherence time.

3.3 Resource configuration in 5G NR

The flexible configuration of 5G time-frequency resources establishes a technical foundation for high-performance transmission of heterogeneous services (Ding et al., 2024).

In the time domain, the 5G frame structure comprises 10 subframes, each with a duration of 1 ms. The number of timeslots within each subframe is determined by the numerology parameter μ ($\mu \geq 0$). Specifically, each subframe is divided into 2^μ timeslots, with each timeslot having a duration of $\tau = 1/2^\mu$ ms. For any value of μ, each timeslot consists of 14 OFDM symbols. 5G supports finer time granularity using mini-slots, which may comprise one or more OFDM symbols. In this work, we schedule URLLC packets at the mini-slot level, where each packet spans a duration of τ_S, referred to as Short TTI (S-TTI). Concurrently, eMBB packets are scheduled at the conventional timeslot level, with each packet spanning a duration of τ_L, denoted as Long TTI (L-TTI). Then, URLLC packets can be transmitted during T_u S-TTIs after generation, where $T_u = D/\tau_S$. Similarly, eMBB packets can be scheduled within each scheduling period over T_e L-TTIs, where $T_e = T/\tau_L$.

In the frequency domain, the available bandwidth W is divided into RBs. The bandwidth of each RB is $w = 12f_b$, where $f_b = 2^\mu \times 15$ kHz is referred to as subcarrier spacing. Benefiting from OFDMA technology in 5G, the orthogonal condition with $\int_0^T \sin(2\pi f_1)\sin(2\pi f_2)d\tau = 0$ can be satisfied between subcarriers, where f_1 and f_2 represent different subcarrier frequencies. This orthogonality enables overlapping spectra among subcarriers without mutual interference. Furthermore, an RU is defined as a resource configuration capable of accommodating the transmission of a packet, which encompasses multiple RBs in the frequency domain and occupies a designated TTI in the time domain. Due to the distinct packet size requirements between URLLC and eMBB services, we define two specialized RU configurations: small-scale RU (S-RU) for URLLC packets and large-scale RU (L-RU) for eMBB packets. We can calculate the number of RBs in an S-RU (L-RU) as $R_S = \lceil b_u/(\eta\tau_S w)\rceil$ ($R_L = \lceil b_e/(\eta\tau_L w)\rceil$). $\lceil\cdot\rceil$ denotes the ceiling operation and η represents the spectral efficiency of the modulation and coding scheme (MCS). Consequently, the bandwidth of an S-RU (L-RU) is $B_S = R_S w$ ($B_L = R_L w$). Let M_S (M_L) be the number of S-RUs (L-RUs) within the total available bandwidth. Then we obtain $M_S = \lfloor W/B_S \rfloor$ ($M_L = \lfloor W/B_L \rfloor$), where $\lfloor\cdot\rfloor$ denotes the floor operation. An example of 5G resource configuration for heterogeneous services with $\mu = 1$ is depicted in Fig. 2.

Unedited

4 Problem formulation for the puncturing method

Because URLLC transmissions cannot wait for the completion of eMBB transmissions, we propose a puncturing method for URLLC traffic over ongoing eMBB traffic. According to the 3GPP Release-15 specification (3GPP, 2018), a PC mechanism is defined for the Physical Uplink Shared Channel. The transmit power of UE is formulated in a simplified notation as follows:

$$P = \min \begin{cases} P_{\max} \\ P_0 + 10\log_{10}(2^{\mu} R_p) + \rho P_l + \Delta_m + f(i) \end{cases}, \tag{2}$$

where $P_{\max}$ represents the maximum transmit power; P_0 denotes the power parameter; R_p indicates the number of RBs allocated to the UE; ρ is the path-loss compensation factor; P_l corresponds to the estimated path-loss between the UE and the BS; Δ_m represents the quality requirement parameter depending on the MCS; and $f(i)$ signifies the parameter for closed-loop PC adjustments. We configure P_0^u and ρ^u for URLLC UEs, while P_0^e and ρ^e for eMBB UEs. By setting $P_0^u \gg P_0^e$, we enable URLLC transmissions to overlay eMBB transmissions effectively. This assumption is justified under the premise that the considered URLLC traffic consists of sporadic packets with an extremely low generation rate (e.g., $\lambda \leq 0.01$), emphasizing URLLC priority while ignoring eMBB-to-URLLC interference.

The specific implementation of the puncturing method is as follows:

- At the beginning of each scheduling period, the BS broadcasts a scheduling list to eMBB UEs, allocating several L-RUs to each eMBB UE. The scheduling list supports dynamic updates across periods in response to real-time variations of eMBB UEs.

- At the start of each S-TTI, each URLLC UE may generate a packet with probability P_a. These URLLC packets then immediately preempt S-RUs within the L-RUs previously assigned to eMBB UEs to transmit. During this puncturing process, URLLC UEs adopt the GF access method and compete for S-RUs within their respective deadline constraints. Because receiver-side decoding is not considered in this scenario, competition will lead to collisions between URLLC UEs. If a packet cannot meet the deadline, it is discarded. To mitigate packet losses caused by collisions and channel errors, the K-repetition mechanism is employed, allowing each packet to compete for S-RUs up to K times. Puncturing intensity is tunable via K for each scheduling period.

Based on the above description, Fig. 3 displays an example of the puncturing method with $N_e = 2$, $N_u = 4$, $K = 4$, and $T_u = 4$.

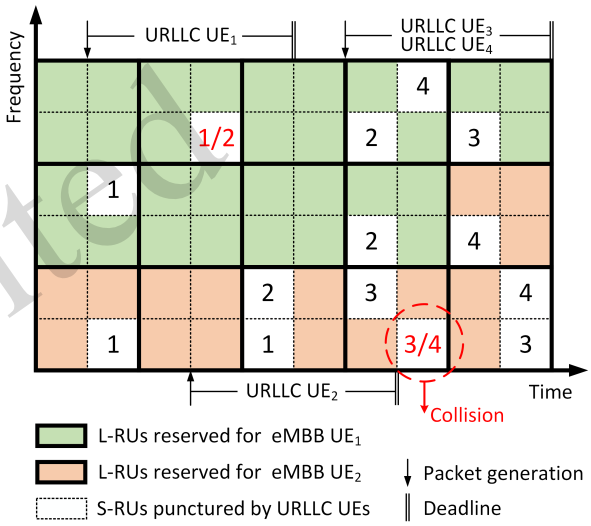

Fig. 3 Example of the puncturing method with $N_e = 2$, $N_u = 4$, $K = 4$, and $T_u = 4$.

Obviously, the puncturing method may cause partial loss of eMBB packets. Because we only focus on the transmission rate of eMBB, the retransmissions of the lost eMBB packets are not taken into account. We employ the coordinates (m, n) ($m \in \{1, 2, \ldots, M_L\}$ and $n \in \{1, 2, \ldots, T_e\}$) to represent the position of an L-RU. Let $r_{m,n}(t)$ denote the number of punctured S-RUs within the $(m, n)^{\text{th}}$ L-RU during period t, where $r_{m,n}(t)$ is closely related to repetition number $K(t)$. Additionally, R represents the total number of S-RUs contained within an L-RU. Then, the transmission rate of the jth ($1 \leq j \leq N_e$) eMBB UE during period t can be approximated as

$$R_e^j(t) = \sum_{m,n} S_{m,n}^j(t) B_L \times \left(1 - \frac{r_{m,n}(K(t))}{R}\right) \log_2(1 + SNR^j(t)). \tag{3}$$

Unedited

$S^j_{m,n}(t)$ is the scheduling parameter for the jth eMBB UE on the $(m,n)^{\text{th}}$ L-RU during period t, that is

$$S^j_{m,n}(t) = \begin{cases} 1 & \text{Case 1 is true} \\ 0 & \text{Case 1 is false} \end{cases}, \tag{4}$$

where Case 1 indicates that the $(m,n)^{\text{th}}$ L-RU is allocated to the jth eMBB UE. $SNR^j(t)$ is the Signal-to-Noise Ratio (SNR) of the jth eMBB UE during period t, expressed as

$$SNR^j(t) = h^j(t)P_e/\sigma^2, \tag{5}$$

where P_e denotes the uplink transmission power of eMBB UEs and σ^2 represents the noise power. $h^j(t)$ is the channel gain of the jth eMBB UE on L-RUs during period t, which follows the Rayleigh or Rician distribution, that is, $h^j(t) = \xi_\beta(t)h^j_0$, where h^j_0 represents the inherent channel gain. Moreover, the intracell interference between eMBB UEs becomes negligible due to the employed OFDMA mechanism.

We formulate the puncturing method as the joint optimization problem **P** to determine $K(t)$ for URLLC UEs and $S^j_{m,n}(t)$ for eMBB UEs in scheduling period t, simultaneously meeting URLLC reliability requirements and maximizing the eMBB accumulated rate:

$$\mathbf{P}\text{: } \underset{S^j_{m,n}(t),K(t)}{\text{Maximize}} \sum_{j=1}^{N_e} R^j_e(t) \tag{6}$$

subject to

$$R_u(K(t)) \geq \delta \tag{6a}$$

$$r_{m,n}(K(t)) \leq R, \forall m,n \tag{6c}$$

$$1 \leq K(t) \leq \hat{K} \tag{6b}$$

$$S^j_{m,n}(t) \in \{0,1\}, \forall m,n,j \tag{6d}$$

$$\sum_{j=1}^{N_e} S^j_{m,n}(t) = 1, \forall m,n \tag{6e}$$

The formulas in Eq. (6) are designed to maximize the accumulated rate of N_e eMBB UEs by optimizing the relevant parameters. Eq. (6a) guarantees that the average reliability of URLLC packets $R_u(K(t))$ meets or exceeds the predefined reliability threshold δ under the corresponding $K(t)$. Eq. (6b) ensures that the number of S-RUs punctured by URLLC within the $(m,n)^{\text{th}}$ L-RU does not exceed the total number of S-RUs in that L-RU. Eq. (6c) specifies the valid range for $K(t)$. Eq. (6d) shows that the scheduling decision of eMBB transmission is binary, as discussed in Eq. (4). Eq. (6e) indicates that each L-RU is allocated exclusively to a single eMBB UE, preventing any overlap or sharing between different eMBB UEs.

5 Reliability model for sporadic URLLC traffic

To enhance resource utilization, we propose a RRCC scheme for sporadic URLLC traffic. Among existing repetition coding schemes, the ORS scheme (Song et al., 2022) allows URLLC UEs with generated packets to immediately select K ($1 \leq K \leq M_S$) RUs in the following single S-TTI based on OFDMA technology in 5G. The Consecutive Repetition Coding-based Contention (CRCC) scheme (Ding et al., 2024) allows URLLC UEs with generated packets to immediately select K RUs in consecutive K ($1 \leq K \leq T_u$) TTIs, while each UE is restricted to accessing only one S-RU per S-TTI. The contention-based repetition scheme (Elayoubi et al., 2019) assumes that all packets are generated at the beginning of the scheduling period, and then each URLLC UE competes for resources K times within the period. The proposed RRCC scheme combines the frequency-domain flexibility of ORS and the time-domain flexibility of CRCC, further expanding the number of available S-RUs. Specifically, after packet generation, the URLLC UE randomly selects K ($1 \leq K \leq \hat{K}$, where $\hat{K}$ denotes the upper limit of K, and $1 \leq \hat{K} \leq T_u M_S$) S-RUs from $T_u M_S$ available S-RUs within the deadline. The example in Fig. 3 also illustrates the principles of the RRCC scheme.

Next, we analyze the transmission reliability $R_u(K)$ of the RRCC scheme. We consider an Observation Window (OW) spanning T_u consecutive S-TTIs, starting from the tth S-TTI and ending at the $(t+T_u-1)^{\text{th}}$ S-TTI. Due to the random nature of the scheme, for each packet generated at the $(t+m)^{\text{th}}$ $(1-T_u \leq m \leq T_u-1)$ S-TTI, there may be repetitions falling within this OW. The expected number of such repetitions at the $(t+m)^{\text{th}}$ S-TTI is $K\frac{T_u-|m|}{T_u}\lambda$. Consequently, the total number of repetitions within the OW is $KT_u\lambda$, which corresponds to $T_u\lambda$ groups of K repetitions. Therefore, the probability that a UE has a group of Krepetitions in T_u consecutive S-TTIs is $P_{\text{a}}(\lambda, T_u) = 1 - e^{-T_u\lambda}$. The transmission of a packet is deemed successful if at least one of the K

Unedited

$$R_u(K) = \sum_{l=1}^{K}(-1)^{l+1}C_K^l(1 - P_a(\lambda, T_u) + P_a(\lambda, T_u)\frac{C_{T_uM_S-l}^K I_{T_uM_S-l\geq K}}{C_{T_uM_S}^K})^{N_u-1}(1-p)^l \quad (7)$$

attempts is successful. Based on the previous preparations, we are now ready to present **Proposition 1**, which gives the analytical expression of $R_u(K)$.

Proposition 1 For a given N_u, T_u, M_S, and p, $R_u(K)$ is formalized in Eq. (7), where I_A is an indicator function whose value is 1 if condition A is true and 0 otherwise.

Proof. Let A_i denote the event that the ith $(1 \leq i \leq K)$ repetition by the UE of interest is successful, then

$$R_u(K) = P\{A_1 \cup A_2 \cup \cdots \cup A_K\}. \quad (8)$$

Given that the contention intensity among UEs remains constant across each S-TTI, we have $P(A_1) = P(A_2) = \cdots = P(A_K)$. Further, we can obtain

$$\begin{aligned} &P\{A_1 \cup A_2 \cup \cdots \cup A_K\} \\ &= \sum_{l=1}^{K}(-1)^{l+1}C_K^l P\{A_1 \cap A_2 \cap \cdots \cap A_l\}, \end{aligned} \quad (9)$$

where $P\{A_1 \cap A_2 \cap \cdots \cap A_l\}$ is the probability that all $l(1 \leq l \leq K)$ repetitions are successful. The calculation of $P\{A_1 \cap A_2 \cap \cdots \cap A_l\}$ is determined by two statistically independent sub-events: (a) the packets from the remaining $N_u - 1$ UEs do not collide with the l repetitions of interest, and (b) all l repetitions are successfully received without channel errors. Thus,

$$P\{A_1 \cap A_2 \cap \cdots \cap A_l\} = P_{\text{nc}}(K, l)^{N_u-1}(1-p)^l, \quad (10)$$

where $P_{\text{nc}}(K, l)$ represents the probability that other $N_u - 1$ UEs do not access the l S-RUs occupied by the UE of interest. The detailed derivation of $P_{\text{nc}}(K, l)$ is presented below. We define $Y_1(K)$ as the number of possible combinations in which a UE selects K S-RUs from the available T_uM_S S-RUs, yielding $Y_1(K) = C_{T_uM_S}^K$. Subsequently, we define $Y_2(K, l)$ as the number of possible combinations where a UE selects K S-RUs from T_uM_S RUs without colliding with the l S-RUs occupied by the UE of interest, resulting in $Y_2(K, l) = C_{T_uM_S-l}^K I_{T_uM_S-l\geq K}$. Then, we have

$$P_{\text{nc}}(K, l) = 1 - P_a(\lambda, T_u) + P_a(\lambda, T_u)\frac{Y_2(K)}{Y_1(K, l)}. \quad (11)$$

By compiling Eqs (8)-(11), we obtain the expression of $R_u(K)$. This completes the proof. □

Using the reliability model in Eq.(7), we can solve the subproblem **P1** in optimization problem **P**, that is,

$$K_{\min} = \arg\min_{1\leq K\leq \hat{K}}(R_u(K) \geq \delta). \quad (12)$$

This subproblem can be efficiently solved through a one-dimensional exhaustive search, which has a computational complexity of $O(n)$.

6 Model-aided DRL solution

6.1 Problem transformation

To adapt to the dynamic environment, we introduce a DRL solution based on a reliability model to effectively solve the puncturing optimization problem. Given that it is difficult to accurately represent the puncturing process using analytical models, we can obtain $r_{m,n}(t)$ for given $K(t)$ through statistical analysis of a specific simulation duration and embed it into the DRL algorithm. Generally, a DRL framework is characterized by its action space A, state space S, and reward function $r(t)$. At each state $s(t) \in S$, the algorithm selects an action $a(t) \in A$ and receives a corresponding reward $r(t)$. Building on this framework, we transform the problem **P** into the following form:

- **State space:** We consider the state space with the channel gains of eMBB UEs, that is, $h^j(t)$, and the packet-generation rate of URLLC UEs, that is, $\lambda(t)$. Therefore, the state during period t can be defined as $s(t) = [h^j(t), \lambda(t)]$ and S is an infinite set.
- **Action space:** The action space we designed consists of two action outputs. The first action is the scheduling list for eMBB, denoted as

$$a_1(t) = \{S_{m,n}^j(t)\colon \sum_{j=1}^{N_e} S_{m,n}^j(t) = 1, \forall m, n\}. \quad (13)$$

Unedited

The second action is the number of repetitions for URLLC, represented as

$$a_2(t) = \{K(t)\colon\ 1 \le K(t) \le \hat{K}\}. \tag{14}$$

Because $a_1(t)$ and $a_2(t)$ are finite sets, we construct a two-output action by pairing these two actions, denoted as $a(t)$. The dimension of the action space can be calculated as $|A| = \hat{K}(N_e+1)^{T_e M_L}$. With the assistance of the reliability model, we can directly calculate the $K(t)$ value that satisfies $R_u(K(t)) \ge \delta$. Thus, $K(t) = K_{\min}(t)$. This effectively reduces the dimension of the action space to $|A| = (N_e+1)^{T_e M_L}$, thereby accelerating training convergence.

- **Reward:** The reward function should quantitatively reflect the requirements of the optimization problem defined in Eq.(6). When the accumulated rate of eMBB UEs increases, a reward should be granted. However, when the reliability requirements are violated, a significant penalty should be imposed. Then, the reward function is designed as

$$r(t) = \frac{c_2}{\lambda(t)}\Big(\sum_{j=1}^{N_e} R_e^j(t) - c_1\zeta\Big), \tag{15}$$

where

$$\zeta = \begin{cases} 0 & \text{if } R_u(K(t)) \ge \delta \\ 1 & \text{otherwise} \end{cases} \tag{16}$$

and c_1 serves as the penalty factor. Furthermore, we introduce $c_2/\lambda(t)$ into the reward function. By adjusting the scaling factor c_2, the reward values for $\lambda(t)$ with varying intensities can be distinguished across different value ranges.

6.2 DDQN algorithm

As a well-established DRL algorithm, the Deep Q-Network (DQN) can solve learning problems in high-dimensional or continuous state spaces by combining the DNN with the Q-learning algorithm. Therefore, it is applicable to the resource optimization problem characterized by a finite action set A and an infinite state set S, as outlined earlier. The Q-value function of DQN is expressed as $Q(s,a;\omega)$, where ω represents the parameters of the neural network. $Q(s,a;\omega)$ is a scalar value representing the

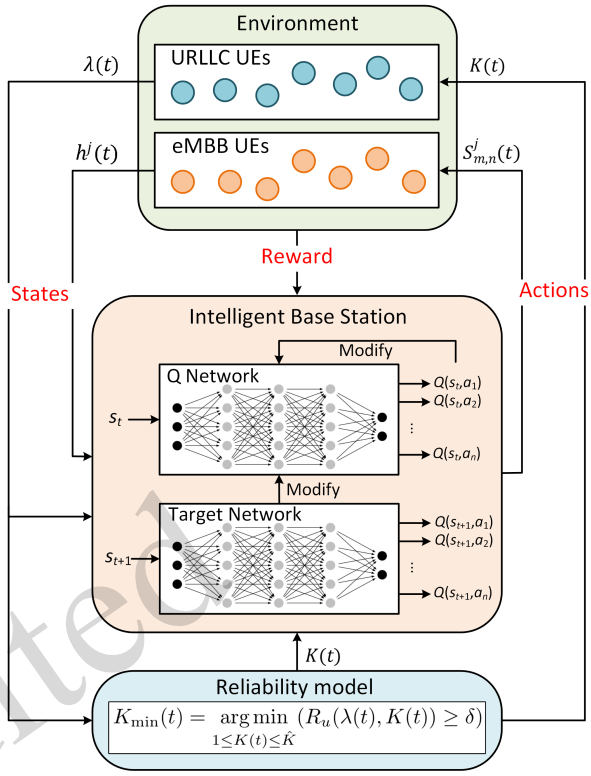

Fig. 4 Model-aided DRL solution with DDQN algorithm for heterogeneous services. DDQN, double DQN; DQN, Deep Q-Network; DRL, Deep Reinforcement Learning.

expected return for taking action a in state s. Therefore, the DQN output is an $|A|$-dimensional vector $\hat{q}$, which contains the value of all actions. The basic DQN process is to randomly initialize ω first, and then iteratively update ω through experience replay. The learning objective is to minimize the difference between the prediction $Q(s,a;\omega)$ and the optimal action-value function $Q^*(s,a)$ for all states s and actions a.

The ϵ-greedy policy is a widely used strategy for balancing exploration and exploitation during experience collection. To mitigate the risk of sub-optimal convergence inherent in purely greedy policies, the ϵ-greedy policy introduces an exploration factor ϵ. Mathematically, the ϵ-greedy policy is defined as

$$a_t = \begin{cases} \arg\max_a Q(s_t,a;\omega) & \text{w.p.} \quad 1-\epsilon \\ \text{a randomly selected } a & \text{w.p.} \quad \epsilon \end{cases}, \tag{17}$$

where the agent explores the action space with probability ϵ.

Due to some common flaws, the DQN is prone to overestimation. Subsequently, the double DQN (DDQN) algorithm (Hasselt et al., 2016) (adopted

Unedited

in our work) is proposed, which introduces a target network $Q(s,a;\omega^-)$ based on the value network $Q(s,a;\omega)$. The value network determines the action with the highest Q-value in the current Q-network, whereas the target network evaluates the value of the action, effectively mitigating overestimation. For the randomly selected (s_x, a_x, r_x, s_{x+1}) from the experience, the DDQN parameters are updated as follows:

$$\omega_{new} \leftarrow \omega_{now} - \alpha \cdot \varphi_x \cdot \nabla_\omega Q(s_x, a_x; \omega_{now}) \quad (18)$$

and

$$\omega^-_{new} \leftarrow \nu \cdot \omega_{new} + (1-\nu) \cdot \omega^-_{now}. \quad (19)$$

In Eq.(18), α represents the learning rate, which controls the step size of parameter updates during optimization. ∇_ω denotes the gradient operator, which is used to compute the direction of steepest ascent in the optimization landscape. φ_x corresponds to the temporal difference (TD) error, defined as $\varphi_x = \hat{q}_x - \tilde{y}_x$, where $\hat{q}_x$ is the predicted Q-value and $\tilde{y}_x$ is the TD target. The TD target, $\tilde{y}_x$, is computed as $\tilde{y}_x = r_x + \gamma \hat{q}_{x+1}$, where r_x is the immediate reward, γ ($\gamma \in [0,1]$) is the discount factor that balances the importance of immediate and future rewards, and $\hat{q}_{x+1} = Q(s_{x+1}, a^*; \omega^-_{now})$ represents the estimated action-value function for the next state. Here, a^* is the action selected according to the ϵ-greedy policy, defined as $a^* = \arg\max_a Q(s_{x+1}, a; \omega_{now})$. In Eq.(19), ν ($\nu \in (0,1)$) is the target smooth factor, a hyperparameter controlling the target network update rate.

In summary, Fig. 4 illustrates the model-aided DRL solution designed for heterogeneous services, which employs the DDQN algorithm.

6.3 Adaptive packet-generation rate

In industrial automation scenarios, certain uplink signals, such as random detection or alarm data, can be characterized as sporadic URLLC traffic. When their detection or alarm frequency is too low (i.e., low packet-generation rate), it may fail to meet industrial control requirements. Conversely, when the detection or alarm frequency is too high (i.e., high packet-generation rate), it may lead to network resource congestion. Thus, our model-aided DRL solution treats $\lambda(t)$ as a variable (Sardar et al., 2024; Elayoubi et al., 2019; Singh et al., 2018) to autonomously adjust the URLLC packet-generation rate according to the target reward. Specifically, we define the target reward interval as $[r_{\min}, r_{\max}]$. By continuously comparing the current reward $r(t)$ with the target reward interval, the system dynamically adjusts $\lambda(t)$ as follows:

- If $r(t) < r_{\min}$, the system reduces $\lambda(t)$ to decrease the URLLC traffic load, that is, $\lambda(t+1) = \max(\lambda(t) - \Delta\lambda, \lambda_{\min})$, where $\Delta\lambda$ is the adjustment step size and $\lambda_{\min}$ is the minimum allowable value for $\lambda(t)$.
- If $r(t) > r_{\max}$, the system increases $\lambda(t)$ to enhance resource utilization, that is, $\lambda(t+1) = \min(\lambda(t) + \Delta\lambda, \lambda_{\max})$, where $\lambda_{\max}$ is the maximum allowable value for $\lambda(t)$.
- If $r(t) \in [r_{\min}, r_{\max}]$, $\lambda(t)$ remains unchanged to maintain system performance within the target reward interval.

While adjusting $\lambda(t)$, the model-aided DRL solution simultaneously optimizes $K(t)$ and $S^j_{m,n}(t)$. Assuming the initial value of $\lambda(t)$ is λ_0, the detailed solving process is provided in **Algorithm 1**.

7 Performance evaluation

In this section, we describe our performance evaluation of the proposed method using experimental validation. For the industrial automation scenarios under consideration, we adopt a subframe-level scheduling framework, with detailed parameter configurations provided in Table 2.

First, we validate the accuracy of the reliability model for the RRCC scheme. Taking $\lambda = 0.005$ and $K = 3$ as an example, we compare the analytical results with the simulation results in Fig. 5, with the simulation duration set to $10^6 T$. As shown, the two sets of results are quite close to the URLLC reliability requirements, demonstrating the feasibility of the derived reliability model.

Then we utilize the reliability model to compare $K_{\min}$ under the dependent variables N_u and T_u among the RRCC, CRCC, and ORS schemes. To facilitate the observation of trends, we set $\delta = 0.9999$, ensuring that each scheme yields a sufficient number of results that meet the reliability requirements. From Fig. 6, we observe that as N_u increases, $K_{\min}$ also gradually increases. Notably, when $N_u > 18$, the ORS scheme fails to meet the reliability requirements. Fig. 7 illustrates that as T_u increases, $K_{\min}$

Unedited

Algorithm 1 Model-aided DRL solution for puncturing method

Input: λ_0, $\Delta\lambda$, $[\lambda_{\min}, \lambda_{\max}]$, $[r_{\min}, r_{\max}]$, h_0^j, system parameters and training parameters
Initialization:
  Set $\lambda \leftarrow \lambda_0$, $h^j \leftarrow h_0^j$
  Initialize DDQN parameters
  Initialize experience replay buffer
for episode $= 1 :$ maxepisodes
  Reset the environment and obtain s
  Set IsDone$\leftarrow 0$
  for step $= 1 :$ maxsteps
    while IsDone $= 0$
      Solve **P1** in Eq.(12) and obtain $K_{\min}(t)$
      Set $a_2 \leftarrow K_{\min}(t)$
      Generate a random $g \in (0, 1)$
      if $g < \epsilon$
        Randomly select a_1 by Eq.(13)
      else
        $a_1 = \arg\max_a Q(s, a; \omega)$
      end
      Calculate r by Eq.(15)
      if $r < r_{\min}$
        $\lambda \leftarrow \max(\lambda - \Delta\lambda, \lambda_{\min})$
        else if $r > r_{\max}$
          $\lambda \leftarrow \min(\lambda + \Delta\lambda, \lambda_{\max})$
        end if
      end if
      IsDone$\leftarrow 0$
      Add $(a, s, r, s'$,IsDone) to replay buffer
      Set $s \leftarrow s'$
      Update DDQN by Eqs(18) and (19)
  end for
end for
Output: Null

for the RRCC scheme gradually decreases. In contrast, the CRCC scheme becomes ineffective when $T_u < 3$, and the ORS scheme does not exhibit performance improvements with increasing T_u. Both Figs 6 and 7 demonstrate that, compared to the other two schemes, the RRCC scheme consistently requires the smallest $K_{\min}$ while satisfying the reliability requirements, thereby minimizing the impact on eMBB during puncturing.

Next, we confirm the convergence of the model-aided DRL solution by comparing it with the model-free DRL solution. The channel fading is modeled as a Rician distribution with $\beta = -50$ dB, which closely approximates a Rayleigh distribution. Subsequently, the channel gain of the jth eMBB UEs is given by $h^j(t) = \xi_{-50\text{dB}}(t)h_0^j$, with $h_0 = [6; 5; 9; 4; 6]$.

Table 2 System parameters

Notation	Value
W	140 MHz
μ	4
τ	0.0625 ms
f_b	240 KHz
w	2.88 MHz
η	1 bits/s/Hz
N_u	2 to 20
b_u	32 bytes
τ_S	0.03125 ms(1 minislot with 7 symbols)
R_S	3
B_S	8.64 MHz
M_S	16
$D(T_u)$	0.25 ms(8 S-TTI)
p	0.001
N_e	5
b_e	1 KB
τ_L	0.25 ms(4 timeslots)
R_L	12
B_L	34.56 MHz
M_L	4
$T(T_e)$	1 ms(4 L-TTI)
P_e	0.1 w
σ^2	1 mw/0 dBm

The parameters $\delta = 0.99999$ and $N_u = 10$ are set. $c_1 = \sum_{j=1}^{N_e} R_e^j(t)$ and $c_2 = 0.01$ are configured in the reward function, where $c_1 = \sum_{j=1}^{N_e} R_e^j(t)$ ensures that the reward becomes 0 if the reliability requirement is not met. Each eMBB UE is allocated at least 2 L-RUs. The statistical duration is set to $3 \times 10^5 T$. The agent parameters and training parameters are detailed in Table 3. In the model-aided DRL, ϵ is set to 0.0004, whereas in the model-free DRL, it is set to 0.0002. The training behavior is shown in Fig. 8. Both solutions exhibit reward convergence over iterations and achieve nearly identical

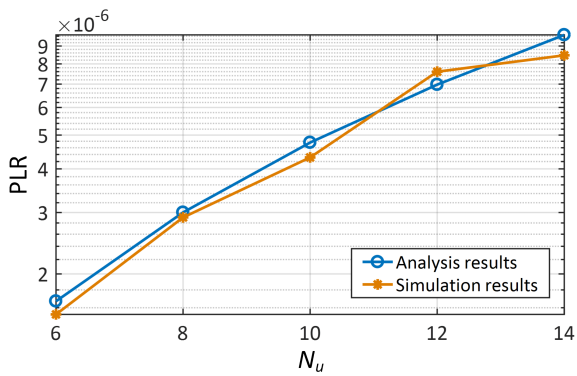

Fig. 5 Accuracy validation of the reliability model for the RRCC scheme with $\lambda = 0.005$ and $K = 3$. RRCC, Random Repetition Coding-based Contention.

Unedited

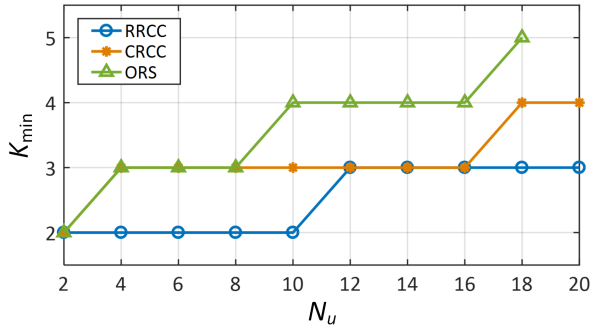

Fig. 6 Comparison of $K_{\min}$ under different N_u among the RRCC, CRCC, and ORS schemes ($\lambda = 0.005$ and $\delta = 0.9999$). ORS, Optimized Repetition Scheme; RRCC, Random Repetition Coding-based Contention.

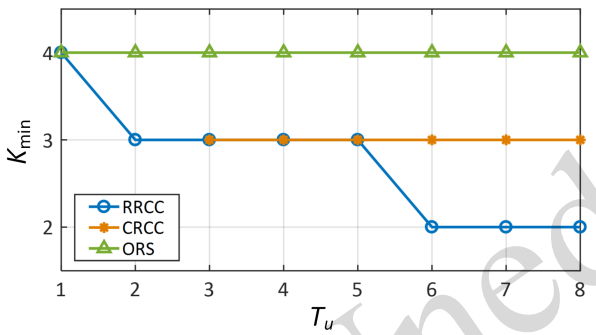

Fig. 7 Comparison of $K_{\min}$ under different T_u among the RRCC, CRCC, and ORS schemes ($\lambda = 0.005$ and $\delta = 0.9999$). ORS, Optimized Repetition Scheme; RRCC, Random Repetition Coding-based Contention.

maximum reward values across different λ. Notably, the model-aided DRL solution achieves convergence in approximately 50% less time than the model-free DRL solution, demonstrating the effectiveness of the reliability model in accelerating convergence.

Table 3 Training parameters

Parameter	Value
Learn rate	0.001
Batch size	5000
Discount factor	0.99
Target smooth factor	0.001
Average window length	100

Moreover, we assess the network performance across various scheduling methods by leveraging the trained agent. The accumulated rate of eMBB UEs, subject to satisfying URLLC reliability requirements, serves as the performance benchmark. A representative channel state vector $h = [6; 5; 9; 4; 6]$ is selected for analysis. Subsequently, we obtain

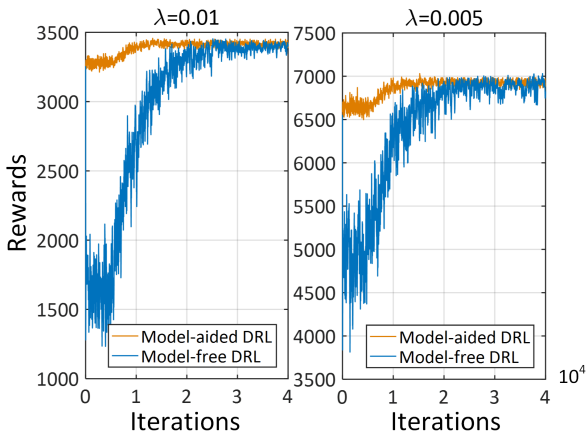

Fig. 8 Reward convergence behavior across training iterations under different λ ($\delta = 0.99999$ and $N_u = 10$).

Fig. 9, which compares the performance of puncturing methods among the ORS scheme, CRCC scheme, and proposed RRCC scheme. The missing data points indicate that the corresponding scheme fails to meet $R_u \leq \delta$ under the given parameters. As observed, the ORS scheme struggles to meet the performance requirements of URLLC traffic as R_u increases. In contrast, both RRCC and CRCC satisfy the URLLC performance requirements, with the RRCC-based puncturing method achieving a higher accumulated rate owing to its fewer repetitions needed (lower K). The overlapping regions between RRCC and CRCC indicate that their K values are identical under the current parameter settings.

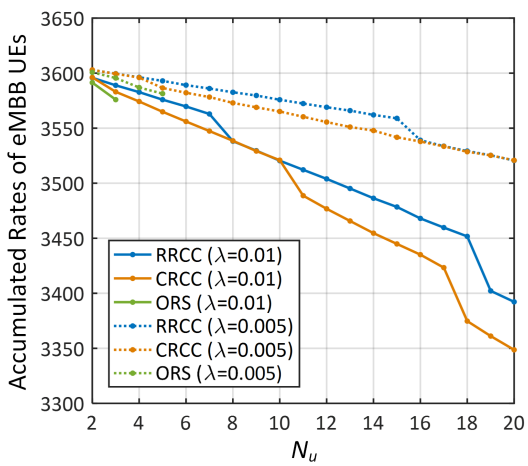

Fig. 9 Performance comparison among puncturing methods using RRCC, CRCC, and ORS schemes under different N_u ($\delta = 99.999\%$). ORS, Optimized Repetition Scheme; RRCC, Random Repetition Coding-based Contention.

Unedited

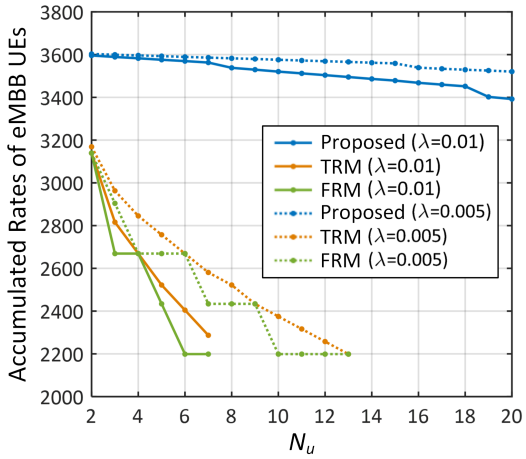

Fig. 10 Performance comparison between the proposed puncturing method and reservation methods under different N_u ($\delta = 99.999\%$).

Furthermore, we compare the proposed puncturing method with two resource reservation methods: (a) the Frequency-domain Reservation Method (FRM) (Ding and Zheng, 2022), where M_1 S-RUs are reserved for URLLC UEs in each S-TTI, and (b) the Time-domain Reservation Method (TRM) (Nomeir et al., 2023), where M_1 S-RUs are reserved for URLLC UEs to transmit accumulated packets at the end of each reservation period. Here, the URLLC deadline is defined as the reservation period. To ensure fairness, URLLC UEs in both the FRM and TRM methods also execute the RRCC scheme proposed in this paper. As shown in Fig. 10, the method proposed in this paper demonstrates significantly superior performance compared to the resource reservation methods due to its enhanced flexibility in resource utilization, and the rigid FRM method exhibits the most severe resource wastage.

In the final stage, we implement adaptive adjustment of the URLLC packet-generation rate. To facilitate adjustment, we define a generation level q, where $q \in \{0, 1, 2, \ldots, 10\}$, and set $\lambda = 10^{-\frac{q}{10}-2}$. We assume $\Delta q = 1$, and q is initialized randomly. Fig. 11 illustrates the adaptive behavior of the trained agent for target reward intervals of $[3300, 3600]$ and $[6900, 7200]$, respectively, under multiple simulation runs. As shown, both sets of results demonstrate that the trained agents are capable of rapidly adjusting to feasible reward values and corresponding λ values.

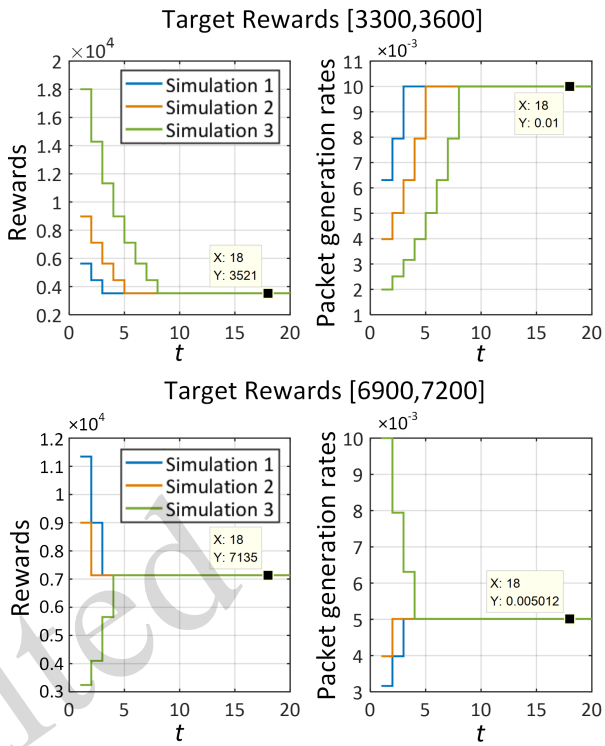

Fig. 11 Adaptive adjustment of the URLLC packet-generation rate across scheduling periods under different target rewards ($N_u = 10$ and $\delta = 99.999\%$). URLLC, Ultra-Reliable Low-Latency Communication.

Unedited

8 Conclusions

In this paper, we describe our investigation of the coexistence problem of URLLC and eMBB services in uplink 5G-based IWNs. Owing to the heterogeneous characteristics, we adopt GF access for sporadic URLLC traffic and SPS for fully buffered eMBB traffic. Then, we propose a puncturing method that allows URLLC transmissions to puncture ongoing eMBB transmissions, formulating an optimization problem to maximize the accumulated rate of eMBB UEs while satisfying the latency and reliability requirements of URLLC UEs. To enhance reliability, we design a RRCC scheme for sporadic URLLC traffic and derive its analytical reliability model. Subsequently, we develop a model-aided DRL solution that jointly optimizes the URLLC repetition number and the eMBB scheduling list, ensuring effective responsiveness to dynamic environments. Using simulation, we demonstrate that the RRCC scheme requires fewer repetitions compared to existing schemes, and the proposed puncturing method achieves higher

resource utilization than the resource reservation method. Meanwhile, the introduced model-aided DRL algorithm is verified to significantly reduce the convergence time.

Contributors

Jingfang DING, Meng ZHENG, and Haibin YU conceived the initial idea and designed the research framework. Jingfang DING and Yitian WANG developed the theoretical proofs and conducted experimental validation. The paper was primarily drafted by Jingfang DING and Meng ZHENG, with Chi XU providing critical review and editorial refinement. All authors contributed to revising the paper and approved the final version.

Conflict of interest

All the authors declare that they have no conflict of interest.

References

3GPP, 2018. Nr; physical layer procedures for data. Release 15. Technical Report No. 3GPP TS 38.214 v15.1.0, https://www.3gpp.org.

Abreu R, Jacobsen T, Pedersen K, et al., 2019. System level analysis of embb and grant-free urllc multiplexing in uplink. IEEE 89th Vehicular Technology Conf, p.1-5. https://doi.org/10.1109/VTCSpring.2019.8746557

Alsenwi M, Tran NH, Bennis M, et al., 2021. Intelligent resource slicing for embb and urllc coexistence in 5g and beyond: A deep reinforcement learning based approach. *IEEE Trans Wirel Commun*, 20(7):4585-4600. https://doi.org/10.1109/TWC.2021.3060514

Ding JF, Zheng M, 2022. Resource allocation and retransmission scheme for urllc in industrial wireless networks with mixed traffic. IEEE 20th Int Conf on Industrial Informatics, p.630-635. https://doi.org/10.1109/INDIN51773.2022.9976174

Ding JF, Zheng M, Wang YT, et al., 2024. A probabilistic repetition coding-based access method for sporadic traffic. *IEEE Trans Veh Technol*, 73(9):14080-14085. https://doi.org/10.1109/TVT.2024.3392643

Elayoubi SE, Brown P, Deghel M, et al., 2019. Radio resource allocation and retransmission schemes for urllc over 5g networks. *IEEE J Sel Areas Commun*, 37(4):896-904. https://doi.org/10.1109/JSAC.2019.2898783

Feng L, Zi YQ, Li WJ, et al., 2020. Dynamic resource allocation with ran slicing and scheduling for urllc and embb hybrid services. *IEEE Access*, 8:34538-34551. https://doi.org/10.1109/ACCESS.2020.2974812

Filali A, Mlika Z, Cherkaoui S, et al., 2022. Dynamic sdn-based radio access network slicing with deep reinforcement learning for urllc and embb services. *IIEEE Trans Netw Sci Eng*, 9(4):2174-2187. https://doi.org/10.1109/TNSE.2022.3157274

Han XH, Xiao K, Liu RQ, et al., 2022. Dynamic resource allocation schemes for embb and urllc services in 5g wireless networks. *Intell Converged Netw*, 3(2):145-160. https://doi.org/10.23919/ICN.2022.0011

van Hasselt H, Guez A, Silver D, 2016. Deep reinforcement learning with double q-learning. Proc Thirtieth AAAI Conf on Artificial Intelligence, p.2094–2100. https://doi.org/10.1609/aaai.v30i1.10295

Huang Y, Li SR, Li CZ, et al., 2020. A deep-reinforcement-learning-based approach to dynamic embb/urllc multiplexing in 5g nr. *IEEE Internet Things J*, 7(7):6439-6456. https://doi.org/10.1109/JIOT.2020.2978692

Ji H, Park S, Yeo J, et al., 2018. Ultra-reliable and low-latency communications in 5g downlink: Physical layer aspects. *IEEE Wirel Commun*, 25(3):124-130. https://doi.org/10.1109/MWC.2018.1700294

Jiang XT, Liang K, Chu XL, et al., 2023. Multiplexing embb and urllc in wireless powered communication networks: A deep reinforcement learning-based approach. *IEEE Wirel Commun Lett*, 12(10):1716-1720. https://doi.org/10.1109/LWC.2023.3290040

Kassab R, Simeone O, Popovski P, 2018. Coexistence of urllc and embb services in the c-ran uplink: An information-theoretic study. IEEE Global Communications Conf, p.1-6. https://doi.org/10.1109/GLOCOM.2018.8647460

Khodakhah F, Stefanovic C, Mahmood A, et al., 2023. Noma or puncturing for uplink embb-urllc coexistence from an aoi perspective? IEEE Global Communications Conf, p.4301-4306. https://doi.org/10.1109/GLOBECOM54140.2023.10437787

Lee K, Kim S, Kim J, et al., 2018. Drama: Device-specific repetition-aided multiple access for ultra-reliable and low-latency communication. IEEE Int Conf on Communications, p.1-6. https://doi.org/10.1109/ICC.2018.8422166

Li J, Zhang X, 2020. Deep reinforcement learning-based joint scheduling of embb and urllc in 5g networks. *IEEE Wirel Commun Lett*, 9(9):1543-1546. https://doi.org/10.1109/LWC.2020.2997036

Liu YW, Clerckx B, Popovski P, 2024. Network slicing for embb, urllc, and mmtc: An uplink rate-splitting multiple access approach. *IEEE Trans Wirel Commun*, 23(3):2140-2152. https://doi.org/10.1109/TWC.2023.3295804

Ma Z, Xiao M, Xiao Y, et al., 2019. High-reliability and low-latency wireless communication for internet of things: Challenges, fundamentals, and enabling technologies. *IEEE Internet Things J*, 6(5):7946-7970. https://doi.org/10.1109/JIOT.2019.2907245

Nomeir MW, Gadallah Y, Seddik KG, 2021. Uplink scheduling for mixed grant-based embb and grant-free urllc traffic in 5g networks. 17th Int Conf on Wireless and Mobile Computing, Networking and Communications, p.187-192. https://doi.org/10.1109/WiMob52687.2021.9606298

Nomeir MW, Gadallah Y, Seddik KG, 2023. Machine learning-based uplink scheduling approaches for mixed traffic in cellular systems. *IEEE Access*, 11:10238-10253. https://doi.org/10.1109/ACCESS.2023.3240569

Popovski P, Trillingsgaard KF, Simeone O, et al., 2018. 5g wireless network slicing for embb, urllc, and mmtc: A communication-theoretic view. *IEEE Access*, 6:55765-55779. https://doi.org/10.1109/ACCESS.2018.2872781

Unedited

Sardar AA, Rao AS, Alpcan T, et al., 2024. Network resource allocation for industry 4.0 with delay and safety constraints. *IEEE Trans Cogn Commun Netw*, 10(1):223-237. https://doi.org/10.1109/TCCN.2023.3319529

Setayesh M, Bahrami S, Wong VWS, 2022. Resource slicing for embb and urllc services in radio access network using hierarchical deep learning. *IEEE Trans Wirel Commun*, 21(11):8950-8966. https://doi.org/10.1109/TWC.2022.3171264

Singh B, Tirkkonen O, Li ZX, et al., 2018. Contention-based access for ultra-reliable low latency uplink transmissions. *IEEE Wirel Commun Lett*, 7(2):182-185. https://doi.org/10.1109/LWC.2017.2763594

Song QJ, She CY, Zheng FC, 2022. Optimization of repetition scheme for urllc with diverse reliability requirements. IEEE 95th Vehicular Technology Conf, p.1-5. https://doi.org/10.1109/VTC2022-Spring54318.2022.9860712

Tian KX, Wang YT, Pan DT, et al., 2024. Drl-based dynamic resource configuration and optimization for b5g network slicing. *IEEE Access*, 12:120864-120876. https://doi.org/10.1109/ACCESS.2024.3452797

Xiao K, Liu X, Han XH, et al., 2020. Flexible multiplexing mechanism for coexistence of urllc and embb services in 5g networks. ITU Kaleidoscope: Industry-Driven Digital Transformation, p.1-9. https://doi.org/10.23919/ITUK50268.2020.9303213

Yuan JT, Yu GD, Yin R, et al., 2020. Group-based data transmission protocol for small-sized urllc services. *IEEE Wirel Commun Lett*, 9(9):1432-1436. https://doi.org/10.1109/LWC.2020.2993194

Zaki-Hindi A, Elayoubi SE, Chahed T, 2020. Urllc and embb coexistence in unlicensed spectrum: a preemptive approach. Int Wireless Communications and Mobile Computing, p.229-234. https://doi.org/10.1109/IWCMC48107.2020.9148222

Zhao LL, Ma SD, Chen WZ, et al., 2023. Semi-probabilistic repetition schemes for sporadic urllc traffic in multiuser massive mimo systems. *IEEE Trans Commun*, 71(4):2104-2120. https://doi.org/10.1109/TCOMM.2023.3241368

Zhou Y, Feng ZY, Song ZQ, et al., 2025. Integrated sensing, communication, and control driven multi-agv closed-loop control. *IEEE Trans Veh Technol*, 74(7):10853-10868. https://doi.org/10.1109/TVT.2025.3546650

Unedited