

# 面向统计时延约束的 Cell-Free 大规模 MIMO 上行传输 UE-AP 连接优化方法

周凡钦, 陈思曼, 丰雷, 杨志祥, 汪旭东, 李文璟

(北京邮电大学网络与交换技术全国重点实验室, 北京 100876)

**摘要:** 为降低信道波动和有限码长传输误差下无蜂窝大规模多输入多输出系统的上行包时延与能耗, 构建了融合顺序门控和解码可靠性的随机服务模型。在到达与服务独立等充分条件下, 基于随机网络演算推导了统计时延界, 并将其作为动态策略的时延参考阈值。以时延超出量和无线发射能耗为联合优化目标, 提出了融合网络演算的近端策略优化算法, 联合选择用户设备与接入点的连接关系、带宽和发射功率。仿真结果表明, 所提算法降低了时延超出量与发射能耗。该方法可为面向统计时延约束的上行连接与资源优化提供参考。

**关键词:** Cell-Free 上行; 随机网络演算; 统计时延约束; 近端策略优化

**中图分类号:** TP393

**文献标志码:** A

## UE-AP Association Optimization with Statistical Delay Constraints in Cell-Free Massive MIMO Uplink Transmission

Zhou Fanqin, Chen Siman, Feng Lei, Yang Zhixiang, Wang Xudong, Li Wenjing

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

**Abstract:** To reduce uplink packet delay and energy consumption under channel fluctuations and finite-blocklength transmission errors, a stochastic service model incorporating ordered gating and decoding reliability was developed for cell-free massive multiple-input multiple-output systems. A statistical delay bound was derived using stochastic network calculus under sufficient conditions including independence between arrivals and service, and was used as a delay reference threshold under dynamic policies. A network calculus aided proximal policy optimization algorithm was proposed to jointly select user equipment–access point associations, bandwidth, and transmit power by minimizing delay excess and wireless transmission energy consumption. Reductions in delay excess and transmission energy consumption were observed in simulations. The proposed method provides a reference for uplink association and resource optimization subject to statistical delay constraints.

**Key words:** cell-free uplink, stochastic network calculus, statistical delay constraints, proximal policy optimization

### 0 引言

随着第五代和第六代移动通信、工业物联网和边缘智能的快速发展, 大量对超可靠低时延通信 (ultra-reliable and low-latency communications, URLLC) 具有严格要求的应用不断涌现, 如工业

自动化控制、车联网协同感知、无人系统编队等, 这类应用通常需要在极短时间内完成数据采集、无线传输与计算决策闭环, 这就对无线链路时延和可靠性提出了远高于传统增强型移动宽带的要求<sup>[1]</sup>。在此背景下, 传统基于小区划分的蜂窝网络逐渐暴露出覆盖不均、边缘用户性能劣化以及频繁切换带

收稿日期: XXXX-XX-XX; 修回日期: XXXX-XX-XX

通信作者: 丰雷, fenglei@bupt.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62571057)

**Foundation Items:** The National Natural Science Foundation of China (No.62571057)

来的时延波动等问题。为应对上述挑战,无线接入架构正由传统蜂窝形态向无蜂窝 (cell-free) 大规模多输入多输出 (multiple-input multiple-output, MIMO) 架构演进,即通过多个分布式接入点 (access point, AP) 协同为用户设备 (user equipment, UE) 提供服务,从而消除小区边界并提升系统整体的可靠性与覆盖一致性<sup>[2]</sup>。

然而,在 Cell-Free 架构下,终端与 AP 之间的连接关系由“单一基站关联”演变为“多 AP 协同服务”,无线侧的连接结构显著复杂化。用户的有效信道质量由多个 AP 的联合接收决定,且受小尺度衰落以及网络负载变化等因素影响而呈现出更强的随机性与时变性<sup>[3]</sup>。在这种情况下,传统基于瞬时信道状态或长期平均速率的连接优化与资源分配方法难以有效刻画无线传输时延的随机波动特性,更难以具有严格时限要求的数据包传输提供可解释可控制的时延上界保障。

随机网络演算 (stochastic network calculus, SNC) 可根据到达与服务过程推导统计时延界和违约概率上界<sup>[4]</sup>。现有分析多面向给定链路或用于性能比较,与在线连接决策的结合仍不充分<sup>[5]</sup>。Cell-Free 上行还需刻画多 AP 联合接收的随机服务特性<sup>[6]</sup>,并将统计时延界引入具有截止时限的数据包调度<sup>[7]</sup>。

本文研究面向边缘计算的 Cell-Free 上行连接优化,在到达与服务过程相互独立、各自增量独立同分布且满足队列稳定性等条件下基于随机网络演算推导统计时延界,并将从到达到成功离队的完整无线侧包时延的超出量作为奖励中的惩罚,联合优化数据包调度、UE - AP 连接和无线资源分配。主要贡献如下:

1) 提出面向统计时延约束的 Cell-Free 上行数据传输模型,用于刻画多用户在共享无线资源下向多个分布式 AP 发送上行数据包的过程,以及同一 UE 缓存队列中数据包到达顺序、等待状态与发送可行性之间的队列依赖关系。在此基础上,建立 UE - AP 连接、带宽分配与发射功率联合优化模型,构建以时延超出量和无线发射能耗最小化为目标的优化问题,并综合考虑包级统计时延、功率分配、带宽占用和 UE - AP 关联等约束;

2) 提出面向有依赖关系的数据包传输的顺序门控随机服务曲线建模方法,将数据包到达过程、

发送使能条件以及多 AP 协同接收下的有限码长解码可靠性统一嵌入随机网络演算框架,用于刻画数据包从到达被调度传输并成功解码的有效服务过程,并进一步推导满足给定违约概率约束的数据包级统计时延上界;

3) 提出面向统计时延约束、融合网络演算的近端策略优化算法 (network calculus aided proximal policy optimization, NCPPO)。该算法将 SNC 计算值作为时延参考阈值,以超出量惩罚嵌入强化学习奖励函数,设计同时包含时延违约惩罚、无线发射能耗和连接决策代价的联合奖励函数。通过对时延违约进行惩罚,引导智能体在随机信道和多用户并发接入场景下自适应优化数据包发送决策、UE - AP 连接关系、带宽占用和功率分配,降低时延超出量与传输能耗,并优化 AP 连接选择;

4) 通过多场景对比实验验证了所提方法的综合优势,在严格时延约束、大规模用户接入、功率受限及不同网络架构下,所提 NCPPO 算法均能显著降低数据包时延违约幅度与无线发射能耗。仿真结果表明,在不同用户规模场景下, NCPPO 相较于原生近端策略优化 (proximal policy optimization, PPO)、随机调度和无调度方案分别可将时延超出量降低约 14.3%、25.0% 和 33.3%,并将无线发射能耗降低约 13.7%、34.8% 和 50.3%;在功率受限场景下, NCPPO 相较于上述基线方案的时延超出量分别降低约 18.4%、31.0% 和 45.9%。相比同类强化学习算法与传统调度方案,所提方法在收敛稳定性、能耗控制和大规模场景扩展性方面表现更优。

## 1 相关工作

在 Cell-Free 大规模 MIMO 上行数据包时延优化领域已有许多相关研究,主要集中在 Cell-Free 上行时延优化和基于随机网络演算的无线时延分析两个方面。

Cell-Free 上行时延优化主要关注时延、能效及资源调度<sup>[8]</sup>,以及时延敏感业务的资源分配<sup>[9]</sup>。孙等人<sup>[10]</sup>联合优化计算卸载、多用户子信道与边缘计算资源; Thai 等人<sup>[11]</sup>采用聚类非正交多址和连续凸近似,优化卸载比例与功率; Tilahun 等人<sup>[12]</sup>利用多智能体强化学习实现分布式通信与计算资源分配。李等人<sup>[13]</sup>研究智能反射面辅助的 Cell-Free 卸

载时延优化, Mahmoudi 等人<sup>[14]</sup>则联合优化联邦学习上行传输的能耗与时延。这些研究为资源优化提供了基础,但包级时延尾部控制及其与在线连接决策的结合仍需加强<sup>[15]</sup>。

SNC 无线时延分析方面,樊等人<sup>[16]</sup>研究多 URLLC 服务的多时间尺度资源分配,但场景限于单小区下行;卢等人<sup>[17]</sup>推导 Cell-Free 上行 URLLC 时延违约概率界,主要针对独立短包的性能分析。李等人<sup>[18]</sup>分析无线衰落和电力线脉冲噪声对混合网络时延的影响;崔等人<sup>[19]</sup>利用 SNC 与矩生成函数推导工业物联网端到端时延界。上述工作尚未充分结合 Cell-Free 上行的多 AP 协同、包级依赖与动态连接决策。

本文进一步考虑随机到达、缓存状态与连接决策之间的包级耦合<sup>[20]</sup>,将顺序门控和有限码长可靠性纳入服务过程,并以 SNC 时延参考阈值指导动态调度。

## 2 系统模型

### 2.1 网络模型

本文考虑一个典型的 Cell-Free 大规模 MIMO 上行系统,由多个用户设备与大量分布式接入点组成。设 UE 集合为  $N = \{1, 2, \dots, N\}$ , AP 集合为  $M = \{1, 2, \dots, M\}$ , 每个 AP 配备  $L_{AP}$  根天线,所有 UE 均为单天线终端。系统工作在时分双工模式下,信道在一个相干区间  $\tau$  内保持不变,相干区间内包含导频传输阶段与上行数据传输阶段,其中导频长度为  $\tau_c$ ,用于信道状态信息的获取。系统总可用带宽为  $B$ 。

各 UE 通过上行链路向 AP 发送数据包,AP 完成解码后将结果转发至中央处理器 (central processing unit, CPU),其中,由于 AP 与 CPU 之间的有线传输链路带宽充足、干扰可控,其传输时延与数据处理时延相对稳定,在后续建模中统一视为常数,从而简化模型复杂度的同时不影响核心特性刻画<sup>[21]</sup>。

Cell-Free 大规模 MIMO 上行系统的网络结构如图 1 所示。

AP - UE 信道采用块衰落模型,块内保持不变、块间相互独立<sup>[22]</sup>。UE  $n$  与 AP  $m$  之间的信道系数为:

$$h_{m,n} = \sqrt{\beta_{m,n}} \zeta_{m,n} \#(1)$$

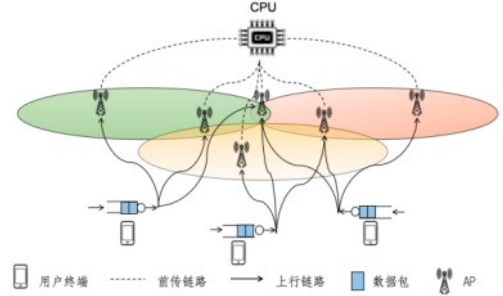


图1 Cell-Free 大规模 MIMO 上行系统

其中,  $\beta_{m,n}$  表示大尺度衰落系数,  $\zeta_{m,n} \sim \mathcal{CN}(0,1)$  表示小尺度瑞利衰落。大尺度衰落建模为:

$$\beta_{m,n} = C d_{m,n}^{-\alpha} z_{m,n} \#(2)$$

其中,  $d_{m,n}$  为 UE 到 AP 的距离,  $\alpha$  为路径损耗指数,  $C$  为参考距离处的衰落常数,  $z_{m,n}$  为对数正态分布的阴影衰落。

### 2.2 数据包顺序依赖模型

每个 UE 独立地生成要发送的数据包,其中第  $n$  个 UE 要发送的数据包队列为  $\mathcal{T}_n = \{1, 2, \dots, T\}$ , 本文规定后一个数据包必须在前一个数据包完成上行传输并被 AP 成功解码后,才允许进入调度或上行传输阶段,即数据包之间存在顺序依赖关系,其可用有向无环图表示为  $G_T = (V_T, E_T)$ , 其中, 顶点集合  $V_T$  表示数据包节点,数据包到达过程服从泊松分布,这一假设与物联网、边缘计算等场景中数据包随机生成的实际特征相符<sup>[23]</sup>; 边集合  $(i,j) \in E_T$  刻画数据包之间的严格先后依赖关系,对于任一数据包  $i$ , 其前置数据包集合为  $P_i$ , 这种依赖关系在实际调度中至关重要,它确保了执行的逻辑正确性。

### 2.3 上行数据包传输过程

在上行数据传输阶段,每个用户向接入点发送数据包。若每个接入点都将所有用户的解码数据传输至 CPU, 会给回程链路带来巨大压力。为解决这一问题,采用以用户为中心的策略: 每个接入点仅需向 CPU 传输其服务用户的检测数据。关联矩阵可筛选出信道条件更优的接入点组成协作集群,避免低质量接入点的干扰与资源浪费,提升协作效率。

为此定义接入点 - 用户关联矩阵  $\mathbf{X} = [x_{m,n}]_{m \in \mathcal{M}, n \in \mathcal{N}}$ , 其中二进制变量  $x_{m,n} \in \{0,1\}$  表示用户  $n$  与接入点  $m$  的关联关系,  $x_{m,n} = 1$  表示接入点

$m$  被选中为用户  $n$  提供服务, 反之则为 0。本文不对 UE - AP 关联施加固定或静态规则, 而是将其建模为可学习的二进制决策变量, 由各 UE 对应的 Actor 在每个调度周期内根据本 UE 的数据包图特征、队列信息及上一时隙链路反馈自适应地选择。

在无线资源分配方面, 假设系统总带宽被划分为  $K$  个互不重叠的带宽份额, 每一份带宽大小为  $B$ , 其中在理想情况下, 每一份带宽仅分配给一个 UE 使用, 以保证用户间的正交传输并避免同频干扰; 然而, 当多个 UE 被调度至同一带宽份额时, 其上行信号将在该频段内相互叠加, 从而引入不可忽略的同频干扰。为刻画带宽资源的分配关系, 定义带宽 - 用户分配矩阵  $\mathbf{Y} = [y_{k,n}] \in \{0,1\}^{K \times N}$ , 其中二进制变量  $y_{k,n} = 1$  表示用户  $n$  被分配至第  $k$  个带宽进行上行传输, 而  $y_{k,n} = 0$  则表示该用户未占用该带宽资源。

与传统蜂窝系统中 UE 仅与服务小区 AP 通信不同, Cell-Free 大规模 MIMO 上行系统中, UE  $n$  的接收信号由空间分布的多个 AP 协同处理, 其瞬时信道质量不再由单一链路决定, 而是由 UE - AP 距离随机变量的空间聚合效应所主导, 为贴合实际部署场景, 本文假设 AP 位置服从二维泊松点过程 (Poisson point process, PPP)  $\Phi_{AP} \sim \text{PPP}(\lambda_{AP})$ , 并考虑 UE 为 Palm 分布下的典型用户。由此, UE  $n$  的上行信干噪比 (signal-to-interference-plus-noise ratio, SINR) 可表示为:

$$\gamma_n = \frac{p_n \sum_{m \in \mathcal{M}} x_{m,n} |h_{m,n}|^2}{\sigma^2 + \sum_{k=1}^K \sum_{n' \neq n} y_{k,n} y_{k,n'} p_{n'}} \quad \#(3)$$

其中,  $p_n$  为 UE 上行发射功率,  $\sigma^2$  为噪声功率。

依据 Polyanskiy - Poor - Verdú 定理<sup>[24]</sup>, 将上述上行瞬时有效 SINR 代入有限码长正态近似, 可以得到 UE  $n$  的条件块解码错误概率近似, 由此推出 Cell-Free 上行距离与衰落联合感知的有限码长解码错误概率定理 (定理 1)。UE 的解码可靠性不再仅由速率选择或瞬时信道条件决定, 而是由 AP 空间密度、UE - AP 距离分布以及传输速率共同塑造, 有限码长解码成功事件随网络空间拓扑和无线信道状态随机变化, 为后续将物理层有限码长可靠性嵌入随机网络演算框架、构建数据包级随机服务过程及统计服务曲线, 并进一步分析数据包统计时

延保障提供了物理层建模基础。定理 1 表述如下:

$$\epsilon_n = Q \left( \frac{\sqrt{\eta\tau} \left( \log_2(1 + \gamma_n) - \frac{r_n}{B} \right)}{\sqrt{V(\gamma_n)}} \right) \quad \#(4)$$

其中,  $\epsilon_n$  表示给定瞬时有效上行 SINR 时, UE  $n$  的条件块解码错误概率,  $B$  为系统带宽,  $r_n$  表示 UE  $n$  在所分配的单个带宽份额上的目标数据传输速率,  $\tau$  为信道块长度,  $\eta = 1 - \tau_c/\tau$  表示有效数据比例,  $V(\gamma) = 1 - (1 + \gamma)^{-2}$  为信道离散度,  $Q(\cdot)$  为高斯  $Q$  函数。

## 2.4 随机网络演算统计时延上界分析

在 SNC 框架下, 不考虑包级依赖时, 上行随机服务的 Mellin 变换为:

$$M_{S_n}(1 - \theta) = E[e^{-\theta S_n}] \quad \#(5)$$

其中,  $S_n$  表示 UE  $n$  的单时隙随机服务量,  $\theta > 0$  为 Mellin 变换中的自由参数。式(5)、(9)~(11)中的变换参数统一记为  $\theta$ 。

同一 UE 内, 后续数据包须等待前序包成功解码后才能发送; 不同 UE 可由多个 AP 并行服务。因此, 有效服务需同时刻画顺序门控、无线信道与有限码长可靠性。

为刻画上述特性, 引入顺序门控变量:

$$G_{n,i}(t) = 1 \{ \forall j \in P_{n,i}, S_{n,j}(t-1) \geq S_{n,j} \} \quad \#(6)$$

其中,  $G_i = 1$  表示数据包  $i$  被允许进入上行传输阶段, 反之则被阻塞。该门控变量用于描述数据包排队处理依赖关系对服务过程的约束作用。

进一步考虑有限码长传输条件下的物理层可靠性, 引入随机变量表示 AP 对数据包的解码是否成功:

$$Z_{n,i}(t) = \begin{cases} 1, & \text{解码成功} \\ 0, & \text{解码失败} \end{cases} \quad \#(7)$$

其中,  $\Pr\{Z_{n,i}(t) = 1\} = 1 - \epsilon_n$ 。

在同时考虑数据包依赖关系与有限码长可靠性的情况下, 用户  $n$  在数据包  $i$  处的有效上行服务过程可表示为:

$$S_{n,i}(t) = G_{n,i}(t) \cdot Z_{n,i}(t) \cdot S_i \quad \#(8)$$

其中,  $S_i$  为数据包  $i$  的大小。该表达式表明, 只有当包级依赖条件满足且物理层解码成功时, 服务过程才对数据包完成产生实际贡献。

在上述建模下, 每个 UE 在一个时隙内仅服务满足顺序条件的队首包。令  $i$  表示时隙  $t$  所服务的队首包, 则 UE 级有效服务量满足  $S_n(t) = S_{n,i}(t) = G_{n,i}(t)Z_{n,i}(t)S_i$ ; 无可服务包时, UE 级有效服务量为 0。式(5)、(9)及后续式(11)中的  $S_n$  统一表示该单时隙服务随机量,  $S_i$  仅表示包长, 其 Mellin 变换为:

$$M_{S_n}(1 - \theta) = E \left[ e^{-\theta G_{n,i}(t) \cdot Z_{n,i}(t) \cdot S_i} \right] \#(9)$$

该顺序门控随机服务曲线模型的关键意义在于: 有限码长传输可靠性通过随机变量  $Z_{n,i}(t)$  显式嵌入服务曲线, 而包级依赖关系则通过门控变量  $G_{n,i}(t)$  对服务过程进行约束。尽管 Cell-Free 系统在物理层采用多 AP 并行接收与合并机制, 同一 UE 内后续数据包仍必须等待前序数据包完成上行传输并被 AP 成功解码后才能获得有效服务。这一建模为后续推导数据包级统计时延上界, 并将其引入连接选择与资源控制决策提供了理论基础。

每个时隙数据包到达量确定的情形下, UE 的到达曲线的 Mellin 变换<sup>[25]</sup>为:

$$M_{A_n}(1 + \theta) = e^{\theta A_n^{\det}} \#(10a)$$

其中,  $A_n^{\det}$  为数据量。由于本文规定数据包到达服从泊松分布, 将泊松随机变量的概率质量函数代入到达过程的 Mellin 变换, 再利用指数函数的幂级数展开可得 UE 的到达曲线的 Mellin 变换为:

$$M_{A_n}(1 + \theta) = \exp \left\{ \lambda_n [e^{\theta S_i} - 1] \right\} \#(10b)$$

其中,  $\lambda_n$  表示 UE  $n$  每时隙的平均到达包数,  $S_i$  表示该固定包长到达模型中的单包大小,  $\theta$  与式(5)、(9)、(10a)一致。

时延违约概率表示时延超过给定界限的概率。对初始为空并按先入先出 (first in first out, FIFO) 顺序服务的基准队列, 采用以下充分条件建立其时延违约概率上界: 累计服务过程由非负单时隙增量累加构成, 并在每条任一路径上满足队列累计离队量的最小加卷积服务下界关系, 从而刻画整个队列的服务能力; 到达增量与服务增量各自在时隙间独立同分布, 两个过程相互独立, 且所需 Mellin 变换有限。在上述条件及下述稳定性条件下, 可得到式(11)所示的时延违约概率上界<sup>[25]</sup>:

$$p_{v,n} \leq \inf_{\theta > 0} \frac{\left( M_{S_n}(1 - \theta) \right)^d}{1 - M_{A_n}(1 + \theta) M_{S_n}(1 - \theta)} \#(11)$$

式(11)在同时满足  $\theta > 0$  及  $M_{A_n}(1 + \theta) M_{S_n}(1 - \theta) < 1$  的稳定参数范围内取下确界。到达与服务过程相互独立使联合指数矩能够分解, 增量在时隙间独立同分布使各区间的变换成为单时隙变换的幂; 对可能形成积压的历史区间求和, 再由上述稳定性条件对几何级数求和, 得到式(11)。由此建立了到达与服务统计量同违约概率上界之间的关系。当式(11)的上界可达到给定容限  $\varepsilon$  时, 选取使该上界不超过  $\varepsilon$  的时延阈值, 即可按式(12)得到对应的统计时延界:

$$D_{n,i}(\varepsilon) = \inf \{ d : p_{v,n} \leq \varepsilon \} \#(12)$$

式(11)(12)所分析的 SNC 虚时延由累计到达与累计成功离队过程定义, 表示从某一观测时刻起, 等待截至该时刻已到达的数据全部成功离队所需的时间。本文将数据包到达 UE 队列作为计时起点, 将其完成上行传输、被 AP 成功解码并离开队列作为计时终点。因此, 用于与统计时延界比较的数据包实际时延采用上述到达至成功离队的完整时间口径, 包含队列等待、前序依赖阻塞、资源等待、无线发送及失败重传所经历的时间。

## 2.5 系统成本

系统成本包括完整无线侧包时延的超时成本和无线发射能耗。时延统计范围为数据包到达 UE 队列至成功完成 UE - AP 上行传输并离队; AP - CPU 回传与处理时延按第 2.1 节视为常数, 不计入与无线队列统计时延界比较的数据包时延<sup>[26]</sup>。记数据包  $i$  从 UE  $n$  发送到 AP 的传输速率为:

$$r_{n,i} = \eta_n B \log_2(1 + \gamma_n) \#(13)$$

在给定传输速率下, 数据包  $i$  的一次无线发送服务时间为:

$$T_{n,i}^{\text{UE-AP}} = \frac{S_i}{r_{n,i}} \#(14)$$

其中,  $\eta_n$  为分配给 UE  $n$  的带宽占比,  $S_i$  为数据包大小。实际时延定义为:

$$T_{n,i}^{\text{arrival}} = a_{n,i} \Delta t \#(15a)$$

$$T_{n,i}^{\text{success}} = s_{n,i} \Delta t \#(15b)$$

$$W_{n,i}^{\text{real}} = T_{n,i}^{\text{success}} - T_{n,i}^{\text{arrival}} = (s_{n,i} - a_{n,i}) \Delta t \#(15c)$$

其中,  $a_{n,i}$  为数据包  $i$  到达 UE  $n$  队列的时隙编号,  $\Delta t$  为一个时隙的持续时间;  $s_{n,i}$  为该数据包成功离队的时隙编号。该时间差包含从到达成功离队期间的队列等待、前序依赖阻塞、资源等待、各次无线发送及失败重传时间; 发生解码失败时,  $s_{n,i}$  不会更新, 计时持续到成功离队。

据此, UE  $n$  的超时成本定义为纳入评价的本 UE 数据包的数据包完整时延超过对应 SNC 时延界的正超出量之和:

$$C_n^t = \sum_{i \in \mathcal{I}_n^t} \max \{0, W_{n,i}^{\text{real}} - D_{n,i}(\epsilon)\} \# (16)$$

其中,  $\mathcal{I}_n^t$  为本轮 UE  $n$  内纳入完整时延评价的已完成数据包集合,  $D_{n,i}(\epsilon)$  为相应时延界。数据包时延与时延界采用相同时间单位; 若统计界以时隙数给出, 应乘以时隙长度后再与物理时间比较。式(16)累加完整包时延的正超出量, 表示超时幅度。

无线发射能耗按实际发射功率与无线发送时间计算, 排队、前序阻塞及资源等待时间不作为无线发射时间。对于一次发送, 其能耗为:

$$C_n^e = p_n \cdot T_{n,i}^{\text{UE-AP}} \# (17)$$

其中,  $p_n$  为该次发送的上行发射功率。存在多次发送或重传时, 累加各次发送能耗; 功率随时隙变化时, 按各时隙实际功率与发射持续时间累加。发射功率由下式确定:

$$p_n = \alpha_n p_n^{\max} \# (18)$$

其中,  $\alpha_n$  为 UE  $n$  的功率控制系数, 即分配给上行数据传输的发射功率占比,  $p_n^{\max}$  为 UE  $n$  的最大上行发射功率。

## 2.6 优化目标及约束条件

本文以功率控制系数  $\alpha$ 、UE - AP 关联关系  $\mathbf{X}$  以及带宽 - 用户分配关系  $\mathbf{Y}$  作为联合优化变量, 优化目标是通过综合考虑多个关键性能指标, 设计一个高效的数据包传输与 AP 连接优化策略  $\pi$ , 以最小化时延超出量  $C^t$  以及无线传输能耗  $C^e$ 。优化目标函数如下:

$$\min_{\alpha, \mathbf{X}, \mathbf{Y}} \left[ (1 - \lambda) C_n^t(\alpha, \mathbf{X}, \mathbf{Y}) + \lambda \sum_i C_n^e(\alpha, \mathbf{X}, \mathbf{Y}) \right] \# (19)$$

$$s.t. C1: S_{n,i}(t) = 0, \text{ if } G_{n,i}(t) = 0 \# (19a)$$

$$C2: \Pr \{ W_{n,i}^{\text{real}} > D_{n,i}(\epsilon) \} \leq \epsilon \# (19b)$$

$$C3: \alpha_n \in [0, 1] \# (19c)$$

$$C4: \sum_{k=1}^K y_{k,n} = 1 \quad \forall n \# (19d)$$

其中,  $\lambda$  是非负加权系数, 用于灵活权衡系统超时程度与能量效率的优化优先级。  $\lambda$  取值越大, 代表优化过程中对能量效率的侧重程度越高, 反之则更注重超时程度的控制。

顺序门控约束 C1 通过定义 UE 级有效服务过程对数据包之间的先后依赖关系进行详细刻画, 当数据包  $i$  的任一前序数据包尚未完成时, 其对应的顺序门控变量为  $G_{n,i}(k) = 0$ , 从而使得该时隙内的有效服务量满足  $S_n(t) = 0$ 。该机制确保即使调度策略在某一时刻选择了尚不满足顺序条件的数据包, 其物理层解码成功事件也不会转化为有效服务; 统计时延约束 C2 要求数据包从到达成功离队的完整时延满足  $\Pr \{ W_{n,i}^{\text{real}} > D_{n,i}(\epsilon) \} \leq \epsilon$ , 该约束描述业务拟达到的概率要求, 超时惩罚用于减小超出幅度; 功率分配约束 C3 引入功率控制系数  $\alpha_n \in [0, 1]$ , 表示用户  $n$  实际用于上行数据传输的发射功率占其最大发射功率的比例, 适当增大比例可提升上行链路的信噪比, 从而提高传输速率、缩短数据包的无线传输时延, 然而过大的功率分配将导致发射能耗显著增加, 并可能加剧系统整体能量开销。因此, 该约束刻画了上行传输中时延性能与能耗之间的基本权衡, 为后续在强化学习框架下实现能耗 - 时延联合优化提供了可行的决策空间; 带宽份额约束 C4 要求每个 UE 只能占用一份带宽。

## 3 基于 PPO 的 UE-AP 连接优化算法

本文将 UE - AP 连接与资源分配建模为非合作随机博弈。每个 UE 具有独立的 Actor、Critic、网络参数及轨迹缓存, 根据本地观测选择资源动作, 并依据自身奖励独立执行 PPO 更新; 所有 UE 动作共同决定系统状态转移。

为将随机网络演算的时延分析用于连接与资源分配, 本文构建融合网络演算的近端策略优化算法 NCPPO。式(11)(12)在初始空队列、FIFO 服务、有效服务下界、到达与服务的独立性及增量独立同分布、Mellin 变换有限性和稳定性等充分条件下, 建立违约概率容限与统计时延界之间的关系。在动态门控与 PPO 控制中, 将其计算值作为式(16)中超时惩罚的参考阈值, 经由各 UE 的奖励进入策略训

练, 形成“时延参考值—超时损失反馈—策略优化”的衔接。

记完整环境状态为  $\mathbf{s}_t$ , 包含各 UE 的数据包到达与队列进度、顺序依赖状态、无线链路状态以及形成上一时隙反馈所需的信息; 必要时将时隙索引纳入状态, 以描述给定联合动作下的 Markov 状态转移。各 UE 不直接获得完整状态, 而仅使用本地观测。UE  $n$  的观测表示为:

$$\mathbf{o}_{n,t} = (\mathbf{z}_n(t), b_n(t), \tilde{d}_n(t), p_n^s(t-1), \gamma_n(t-1)) \# (20)$$

其中,  $\mathbf{z}_n(t)$  为有向图卷积网络 (directed graph convolutional network, DGCN) 提取的本 UE 数据包依赖图嵌入;  $b_n(t)$  为队首任务就绪标志;  $\tilde{d}_n(t)$  为归一化剩余数据量;  $p_n^s(t-1) = 1 - \epsilon_n(t-1)$  和  $\gamma_n(t-1)$  分别为上一时隙的解码成功概率与 SINR。该观测是完整环境状态的局部表征, 不要求其自身满足 Markov 性。

在每个时隙  $t$ , 各 UE 的 Actor 分别根据本地观测输出自身动作, 环境将这些动作汇集为联合动作:

$$\mathbf{a}(t) = [a_1(t), a_2(t), \dots, a_N(t)] \# (21)$$

其中,  $\mathbf{a}_n(t) = (\alpha_n(t), \mathbf{x}_n(t), \mathbf{y}_n(t))$  为 UE  $n$  的资源动作;  $\mathbf{x}_n$ 、 $\mathbf{y}_n$  分别为关联矩阵  $\mathbf{X}$ 、分配矩阵  $\mathbf{Y}$  中对应 UE  $n$  的列。采用功率集合  $P = \{0.2, 0.5, 0.8\}$  作有限动作近似<sup>[27]</sup>, 并非连续功率采样, 记所有候选 AP 连接组合构成的集合为  $\Omega_{AP}$ , 其中每个元素均为一个非空 AP 子集。令  $M_c = \min(2, M)$  为每个 UE 允许连接的 AP 数量上限, 则  $\Omega_{AP} = \{\omega \subseteq \{1, \dots, M\} : 1 \leq |\omega| \leq M_c\}$ , 其中  $\omega$  表示一个候选 AP 子集,  $|\omega|$  表示该子集中的 AP 数量。因此, 每个 UE 可选择连接一个 AP, 或在  $M \geq 2$  时选择连接两个不同的 AP。每个 UE 的候选动作集合为:

$$\mathcal{A}_n = P \times \Omega_{AP} \times \{1, \dots, K\} \# (21a)$$

每个类别索引  $j$  唯一对应三元组  $(\alpha_j, \omega_j, k_j)$ , 其中  $\omega_j \in \Omega_{AP}$  为该类别选中的 AP 子集; 式(21b)中  $m \in \omega_j$  表示 AP  $m$  属于该子集, 即 UE  $n$  与 AP  $m$  建立连接。采样后按下式确定发射功率及二进制矩阵元素:

$$p_n = \alpha_j p_n^{\max}, x_{m,n} = 1 \{m \in \omega_j\},$$

$$y_{k,n} = 1 \{k = k_j\} \# (21b)$$

非合作随机博弈中的各决策主体分别追求自身

长期期望收益, 其收益及下一状态同时取决于各主体动作和随机因素<sup>[28]</sup>。本文各 UE 通过共享带宽上的上行干扰发生交互: 一个 UE 提高功率可改善自身 SINR, 同时增加同频 UE 所受干扰; 带宽选择改变同频干扰关系, AP 连接选择影响接收增益与所受干扰。各 UE 动作共同影响解码成功概率、有效服务量及后续队列状态, 有限码长解码成败又使状态转移具有随机性, 因此, 个体目标、动作耦合和随机状态演化构成该博弈建模的依据。UE  $n$  的即时奖励为:

$$R_n(a_n, a_{-n}) = -(1 - \lambda) C_n^t(a_n, a_{-n}) - \lambda C_n^c(a_n, a_{-n}) \# (22)$$

其中,  $a_n$ 、 $a_{-n}$  分别表示 UE  $n$  的动作及其他 UE 的动作集合,  $\lambda$  用于权衡时延成本与能耗。记时隙  $t$  的个体奖励为  $r_{n,t} = R_n(a_n(t), a_{-n}(t))$ , 各 UE 仅使用自身奖励更新; 系统平均奖励仅用于性能统计与展示。

式(22)中的时延成本以式(15a) - (15c)定义的完整无线侧包时延为评价对象。对每个数据包保存其到达时刻及对应时延参考阈值, 计时从到达开始, 在成功离队时停止。逐时隙奖励对已到达且尚未完成的数据包 (包括当前时隙完成的数据包), 计算自到达起的已逝时间超过该阈值的正超出量在本时隙内的增量, 并将各包增量之和作为 UE  $n$  当前时隙的时延成本。对时延界有限、且界值在包存续期间保持不变的已完成数据包, 上述增量从到达到完成的累计值等于式(16)中的完整时延正超出量; 包成功离队后不再增加其惩罚。UE - AP 关联、功率及带宽分配通过改变链路质量、服务速率与解码成功概率, 影响包的服务进度、等待时间及最终成功离队时刻, 进而改变超时成本。由此, 奖励与统计时延分析均围绕到达至成功离队的时间范围进行比较, 同时由能耗项评价实际无线发射开销。

每个 UE 的 Actor 接收式(20)定义的本地观测  $\mathbf{o}_{n,t}$ , 对  $J$  个资源组合输出分数  $l_{n,j}$ , 并通过 Softmax 参数化一个 Categorical 分布<sup>[27]</sup>:

$$q_{\theta_n}(j|\mathbf{o}_n) = \frac{\exp(l_{n,j})}{\sum_{v=1}^J \exp(l_{n,v})} \# (23a)$$

由该分布采样  $j_n$  并映射为  $a_n$ ; 功率档位、AP 子集和带宽编号在同一个类别中联合选择, 不对三者作独立性假设。给定各 UE 观测, 采样随机性相

互独立, 系统联合策略为:

$$\pi_{\Theta}(a|s) = \prod_{n=1}^N q_{\theta_n}(j_n|o_n) \# (23b)$$

其中,  $\Theta$  表示各 UE 策略参数的集合,  $o_n$  由系统状态  $s$  提取; 式(23b)的乘积分解仅表示给定观测下的独立采样, 不表示 UE 间无干扰耦合。动作集合预先排除空 AP 集、重复 AP 及超出连接上限的组合, 并保证每个 UE 恰选一份带宽; 不同 UE 可复用同一份带宽并产生干扰。因此当前不设置额外的状态相关动作掩码, 也不对采样动作作事后裁剪或可行性投影。包级顺序约束由队列门控执行, 仅服务已到达且前序完成的队首包; 无可服务包时该 UE 不产生有效服务。

每个 UE 的 Critic 以本地观测为输入,  $V_{\phi_n}(o_{n,t})$  近似估计当前各 UE 策略共同作用下, 该 UE 未来折扣奖励之和的条件期望。各 UE 分别训练, 不共享参数; 该本地价值估计及独立更新方式属于独立 PPO<sup>[29]</sup>。记全部 UE 策略为  $\pi_{\Theta}$ , UE  $n$  的条件期望回报定义为:

$$V_n^{\pi_{\Theta}}(o_{n,t}) = E_{\pi_{\Theta}}[\sum_{k=0}^{\infty} \gamma^k r_{n,t+k} | o_{n,t}] \# (24)$$

式(24)的期望依赖全部 UE 策略, 其他 UE 的策略更新会改变本 UE 面对的环境。局部 Critic 用于近似该条件期望, 不将局部观测视为完整的 Markov 状态。记式(25)中的  $Q$  为在联合策略  $\pi_{\Theta}$  下, 进一步给定本 UE 观测及动作时的条件期望回报, 其余 UE 动作按各自策略产生, 对应优势为:

$$A_{n,t}^{\pi_{\Theta}} = Q_n^{\pi_{\Theta}}(o_{n,t}, a_{n,t}) - V_n^{\pi_{\Theta}}(o_{n,t}) \# (25)$$

Critic 估计本 UE 的长期回报。由于条件期望回报难以精确计算, 实际训练使用本 UE 的奖励与局部价值估计构造时序差分 (temporal difference, TD) 误差:

$$\delta_{n,t} = r_{n,t} + \gamma V_{\phi_{n,old}}(o_{n,t+1}) - V_{\phi_{n,old}}(o_{n,t}) \# (26)$$

对采样轨迹, 进一步采用广义优势估计 (generalized advantage estimation, GAE):

$$\hat{A}_{n,t} = \sum_{l=0}^{H-t-1} (\gamma \lambda_{GAE})^l \delta_{n,t+l} \# (27)$$

其中,  $\gamma$  为折扣因子,  $\lambda_{GAE} \in [0, 1]$  为 GAE 参数, 与式(22)中的成本权重  $\lambda$  区分;  $H$  为本段轨迹结束时刻。  $V_{\phi_{n,old}}$  表示采样时的 Critic, 真实终止状态的自举价值置零, 采样截断时采用末状态的价值估计。上述局部 TD 误差和 GAE 作为训练近似, 用于权衡

优势估计的偏差与方差<sup>[30]</sup>。

SNC 通过各 UE 自身的奖励进入优势估计: 式(26)中的  $r_{n,t}$  包含本 UE 超时惩罚, 其变化影响 TD 误差及式(27)的 GAE。各 Critic 据此拟合本 UE 回报, 各 Actor 则按式(29)更新本 UE 动作概率, 形成“统计时延界一个体奖励—本 UE 优势估计—独立策略更新”的反馈过程。

对 UE  $n$ , 记录采样时旧策略对类别  $j_{n,t}$  的对数概率, 在同一观测、同一类别索引下计算新策略概率。由 PPO 的新旧策略比定义<sup>[31]</sup>, 更新时采用:

$$\rho_{n,t} = \frac{q_{\theta_n}(j_{n,t}|o_{n,t})}{q_{\theta_{n,old}}(j_{n,t}|o_{n,t})} \# (28)$$

由于类别到资源组合一一对应, 无须连续密度的 Jacobian 修正或投影后的概率重算。式(23b)对应的联合动作概率比在数学上等于各 UE 概率比的乘积, 本文实际按 UE 独立裁剪:

$$L_n^{\text{PPO}}(\theta_n) = \hat{E}_t[\min\{\rho_{n,t} \hat{A}_{n,t}, \text{clip}(\rho_{n,t}, 1 - \delta, 1 + \delta) \hat{A}_{n,t}\}] \# (29)$$

其中,  $\delta = 0.2$  为 PPO 裁剪参数, 与时延违约概率容限  $\varepsilon$  不同;  $\hat{A}_{n,t}$  由 UE  $n$  自身的时序差分误差和 GAE 计算。

每个 UE 的价值网络分别采用均方差损失:

$$L_n^V(\phi_n) = \hat{E}_t[(V_{\phi_n}(o_{n,t}) - \hat{G}_{n,t})^2] \# (30)$$

其中,  $\hat{G}_{n,t} = \hat{A}_{n,t} + V_{\phi_{n,old}}(o_{n,t})$  为采样轨迹构造的本 UE 回报目标, 在本轮更新中固定<sup>[31]</sup>。每个 UE 分别优化自身的 Actor 与 Critic, 损失为:

$$L_n = -L_n^{\text{PPO}}(\theta_n) + c L_n^V(\phi_n) \# (31)$$

其中,  $c$  为价值损失权重。各 UE 仅更新自身参数, 式(31)并非使用全局奖励联合更新全部 UE 的共享网络。

训练时, 各 UE 独立采样动作, 环境统一计算同频干扰、SINR、解码结果和门控后的有效服务, 推进队列状态, 并向各 UE 反馈各自的下一观测与奖励。各 UE 分别缓存轨迹、计算 GAE 并执行 PPO 更新, 流程见算法 1。

奖励计算以式(12)给出的时延参考阈值为比较基准, 采用完整无线侧包时延的正超出量增量构造各 UE 的逐时隙时延成本, 并按实际发射持续时间计算能耗。

**算法 1** NCPPO 算法**输入** 各 UE 初始参数  $\{\theta_n, \phi_n\}$ , 最大迭代次数 $K_{\max}$ , 轨迹长度  $H$  及梯度更新次数  $U$ **输出** 各 UE 训练后的参数  $\{\theta_n^*, \phi_n^*\}$ 初始化各 UE 独立的 Actor - Critic、优化器及轨迹缓存  $B_n$ , 令迭代计数  $k = 0$ 1) while  $k < K_{\max}$  do2)  $k \leftarrow k + 1$ ; 重置环境及各 UE 缓存  $B_n$ , 保存采样策略与价值网络参数;3) for 每个环境交互时隙  $t$  do

4) 各 UE 按式(20)取得本地观测, 按式(23a)独立采样类别, 并依式(21b)映射为自身资源动作;

5) 环境执行各 UE 动作组成的联合动作, 计算干扰、解码及门控服务, 更新队列, 返回各 UE 的  $o_{n,t+1}$  与  $r_{n,t}$ ; 奖励按式(16)、(17)、(22)分别计算;6) 各 UE 将自身观测、类别索引、旧策略对数概率、价值估计、奖励及终止标志分别写入  $B_n$ ;

7) end for

8) for 每个 UE 及其缓存中的时间步 do

9) 按式(26)计算本 UE 的 TD 误差  $\delta_{n,t}$ ;10) 按式(27)计算本 UE 的 GAE 估计  $\hat{A}_{n,t}$ ;11) 计算回报目标  $\hat{G}_{n,t} = \hat{A}_{n,t} + V_{\phi_{n,\text{old}}}(\mathbf{o}_{n,t})$ ;

12) end for

13) for 各 UE 及其  $U$  次梯度更新 do14) 从本 UE 缓存取样, 按式(28)计算同一类别的新旧策略概率比  $\rho_{n,t}$ ;

15) 按式(29)计算本 UE 的 PPO 裁剪目标;

16) 按式(30)计算本 UE 的价值损失;

17) 按式(31)更新本 UE 的 Actor 参数  $\theta_n$  和 Critic 参数  $\phi_n$ ;

18) 保持其他 UE 参数不变;

19) end for

20) end while

计算复杂度分析: 在固定数据包图规模及网络结构下, 记  $P_{\text{tot}}$  为全部 UE 网络的参数量之和,  $N$  为 UE 数,  $H$  为单轮采样时隙数,  $U$  为每轮每个 UE 的梯度更新次数,  $B_{\text{tr}}$  为每次更新使用的本 UE 轨迹数据量。按网络计算量与参数量成正比的近似, 策略采样开销为  $O(HP_{\text{tot}})$ , 各 UE 的 TD 误差与 GAE 计算开销为  $O(NH)$ , 更新开销为  $O(UB_{\text{tr}}P_{\text{tot}})$ ,  $K_{\max}$  轮的网络训练总开销为  $O(K_{\max}(HP_{\text{tot}} + NH +$

$UB_{\text{tr}}P_{\text{tot}}))$ 。该估计不包含无线环境仿真和 SNC 时延界计算的开销。

## 4 仿真与分析

### 4.1 参数设置

本文的主要仿真参数如表 1 所示。

| 表 1 主要仿真参数 |           |           |      |
|------------|-----------|-----------|------|
| 参数         | 说明        | 取值        | 单位   |
| $L$        | 覆盖区域边长    | 200       | m    |
| $\alpha$   | 路径损耗指数    | 3.0       | —    |
| $C$        | 路径损耗常数    | 1.0       | —    |
| $\sigma^2$ | 热噪声功率     | $10^{-9}$ | W    |
| $p^{\max}$ | UE 最大发射功率 | 0.1       | W    |
| $R_0$      | 单份资源基准速率  | 300       | Mbps |
| $\Delta t$ | 时隙长度      | 1.0       | ms   |
| $\tau_c$   | 导频长度      | 10        | 符号   |
| $\tau$     | 相干块长度     | 200       | 符号   |
| $S_i$      | 数据包大小     | 0.1 - 1.0 | MB   |

### 4.2 仿真分析

对比方法包括原生 PPO、融合网络演算的异步优势演员 - 评论家 (network calculus asynchronous advantage actor-critic, NCA3C)<sup>[32]</sup>、随机调度和无调度。原生 PPO 用于比较引入 SNC 时延惩罚的效果, NCA3C 用于比较不同策略更新机制; 随机调度随机分配资源, 无调度采用全带宽并发传输。实验从超时程度、发射能耗、网络架构、用户规模和功率约束等方面评估 NCPPO。

图 2 中, 三种算法初期奖励较低且波动明显。NCPPO 约在 2000 步后快速收敛, PPO 和 NCA3C

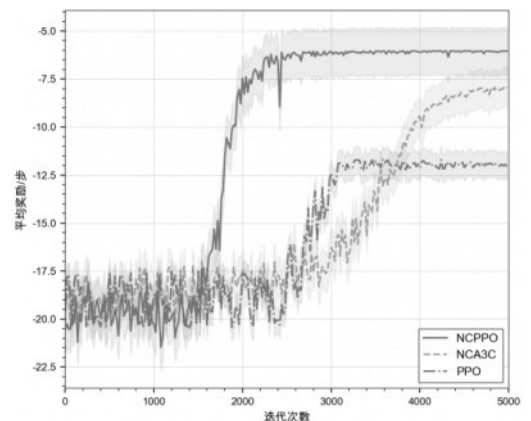


图 2 不同算法奖励情况

分别约在 3000 步和 4000 步后趋于稳定。初期波动与随机探索有关；NCPPO 利用 SNC 时延基准构造奖励，使策略获得超时程度的反馈。其稳定阶段的表现与时延评价和 PPO 裁剪更新相结合的设计相一致。稳定阶段，NCPPO 奖励更高、波动更小，表现出较好的收敛速度与稳定性。

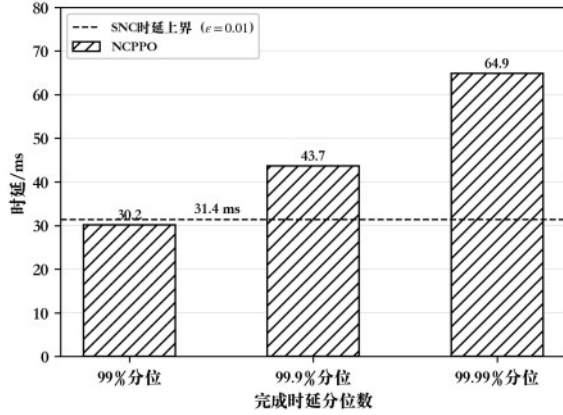


图3  $\varepsilon=0.01$  时 SNC 时延上界与完成时延分位数的对比示意

图 3 中，SNC 在  $\varepsilon=0.01$  时给出的时延上界为 31.4 ms，三个分位时延依次为 30.2 ms、43.7 ms 和 64.9 ms。时延上界高于 99% 分位时延 1.2 ms，说明该阈值能够覆盖至少 99% 的数据包，满足  $\varepsilon=0.01$  所对应的违约概率不超过 1% 的统计时延约束要求，可表明时延上界在该场景下的保守性和有效性提供经验支持，进一步支撑允许非零违约概率的统计时延保障结论。

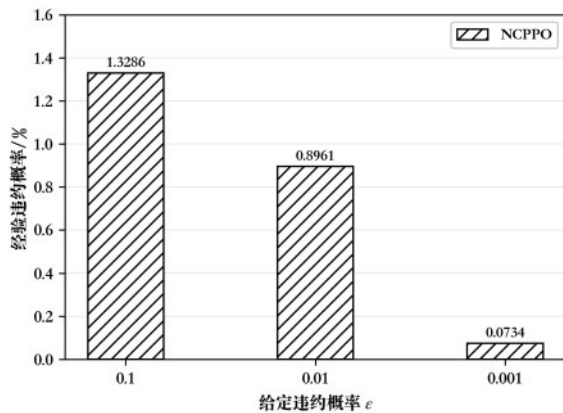


图4 不同  $\varepsilon$  下 NCPPO 的经验违约概率

图 4 给出了 NCPPO 在不同给定违约概率容限  $\varepsilon$  下的经验违约概率。当  $\varepsilon$  分别为 0.1、0.01 和 0.001 时，数据包完整无线侧时延超过对应实际违约概率

分别为 1.3286%、0.8961% 和 0.0734%，均低于对应的 10%、1% 和 0.1%，三种时延超界比例均满足相应要求。

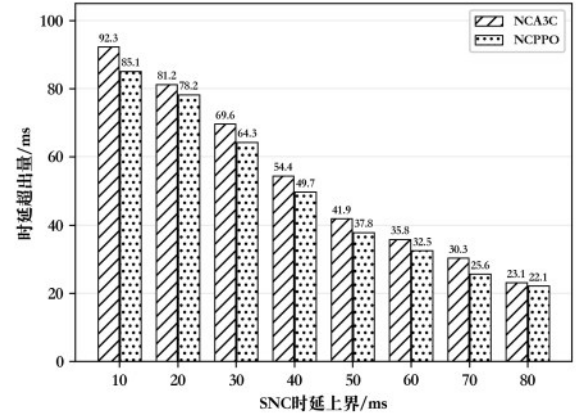


图5 不同算法在不同 SNC 时延上界下的数据包超时程度对比

图 5 中，时延界由 10 ms 放宽至 80 ms 时，两种算法的超时程度均下降，NCPPO 始终较低。时延界为 10 ms 时，NCA3C 与 NCPPO 的超出量分别为 92.3 ms 和 85.1 ms，后者降低约 7.8%。时延界放宽会直接减小式(16)中的正超出量，因此两条曲线整体下降。相同时延界下，NCPPO 的超出量更小，说明其学得连接与资源分配策略具有较好的超时控制效果。

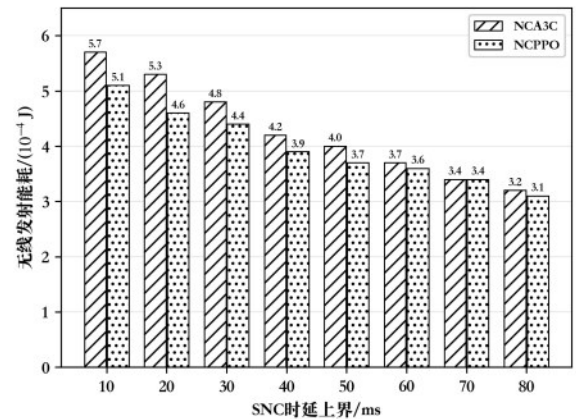


图6 不同算法在不同 SNC 时延上界下的无线发射能耗对比

图 6 中，两种算法的发射能耗均随时延界放宽而下降。时延界为 10 ms 时，NCA3C 与 NCPPO 能耗分别为  $5.7 \times 10^{-4}$  J 和  $5.1 \times 10^{-4}$  J，后者降低约 10.5%，体现了较好的时延 - 能耗权衡。时延界放宽减轻了超时惩罚，为策略降低发射功率留出空间；能耗项则持续约束发射开销。NCPPO 在相同

时延界下保持较低能耗,说明其能够利用连接与资源分配改善这一权衡。

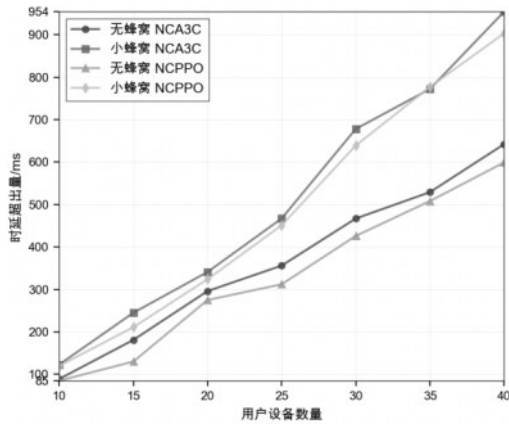


图7 不同算法在 Cell-Free 与 Small Cell 架构下随用户设备数量变化的数据超时程度对比

图7中, Cell-Free 架构及 NCPPO 均有助于降低超时程度。UE 数量为40时, Cell-Free 下 NCPPO 的时延超出量较 NCA3C 低约 50 ms; 多 AP 协作提供了连接与资源协调空间。随着用户增加, 共享频谱上的干扰与资源竞争更突出。Cell-Free 可利用不同 AP 组合改善有效链路质量, 而 NCPPO 依据超时反馈调整连接与资源分配, 两者结合有助于缓解用户规模增长带来的时延压力。

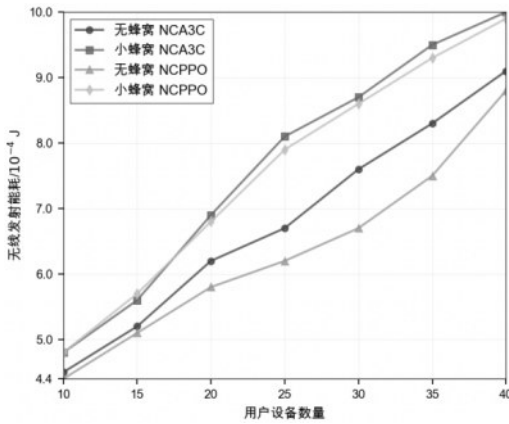


图8 不同算法在 Cell-Free 与 Small Cell 架构下随用户设备数量变化的无线发射能耗对比

图8中, NCPPO 与 Cell-Free 组合的能耗最低; 随着用户增加, Small Cell 的能耗增长更快, 反映出多 AP 协作的能耗优势。多 AP 联合服务能够提高所选链路的有效接收增益, 为降低发射功率提供条件; Small Cell 的连接选择空间较小, 用户增多时更易受到同频干扰影响。因此, 联合优化连接与功

率有助于控制能耗增长。

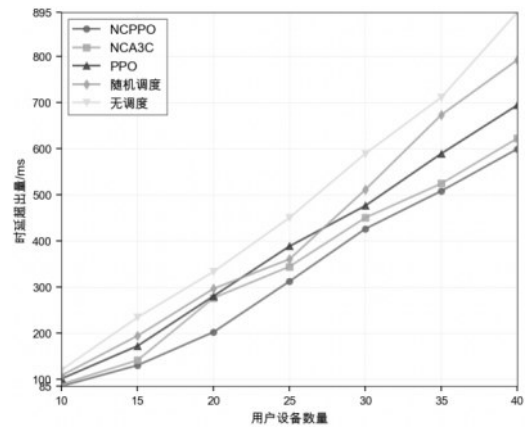


图9 不同调度策略随用户设备数量变化的数据超时程度对比

图9中, UE 数量为40时, NCPPO 较 NCA3C、原生 PPO、随机调度和无调度的时延超出量分别降低约 3.2%、14.3%、25.0% 和 33.3%。NCPPO 与 NCA3C 均通过超时惩罚获得时延界相关反馈; 随机调度和无调度缺少相应的自适应调整, 用户增加时更难协调资源竞争。各方法间的差异体现了时延评价与动态调度共同作用的价值。结果表明, SNC 时延评价与策略优化的结合有助于抑制超时程度。

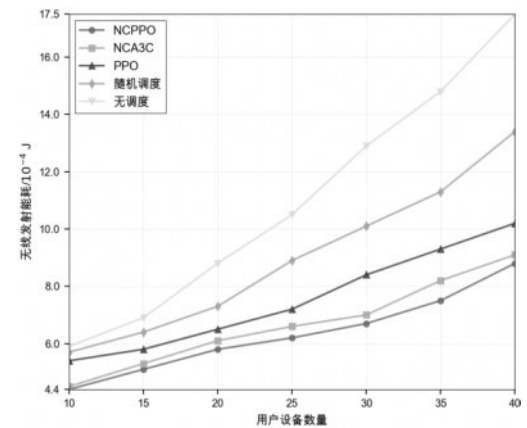


图10 不同调度策略随用户设备数量变化的无线发射能耗对比

图10中, UE 数量为40时, NCPPO 较 NCA3C、原生 PPO、随机调度和无调度的能耗分别降低约 3.3%、13.7%、34.8% 和 50.3%。奖励中的能耗项使策略在改善超时程度时同时评价功率开销, AP 选择与带宽协调又可减少对提高功率的依赖。因此, NCPPO 的能耗优势体现了连接和资源联合优化的作用。上述结果说明, 其在降低超时程度的同时兼顾了发射能耗。

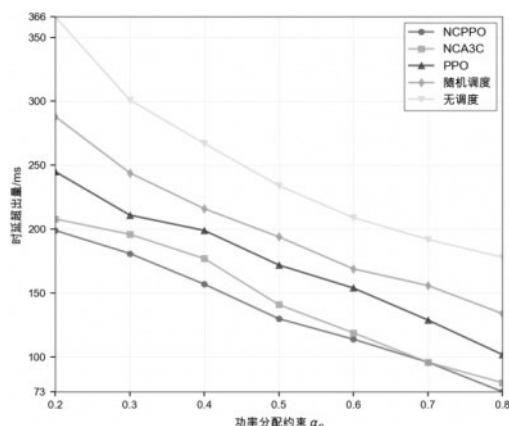


图11 不同算法在不同功率分配约束下的数据包超时程度对比

图11中,在最严格功率约束下,NCPPO较NCA3C、原生PPO、随机调度和无调度的时延超出量分别降低约4.8%、18.4%、31.0%和45.9%。功率上限收紧后,单纯提高发射功率的调节空间缩小,AP选择和带宽协调更为关键。NCPPO利用超时反馈优化这些决策,因此在功率受限场景下仍能获得较小的时延超出量。上述结果表明,连接与带宽优化在功率受限时仍然有效。

图12中,在最宽松功率约束下,NCPPO较NCA3C、原生PPO、随机调度和无调度的能耗分别降低约2.2%、12.0%、18.5%和26.7%。功率上限放宽仅扩大可选功率范围,并不要求始终采用高功率。NCPPO的能耗惩罚持续评价额外功率开销,并结合AP选择与带宽分配调整传输策略,从而抑

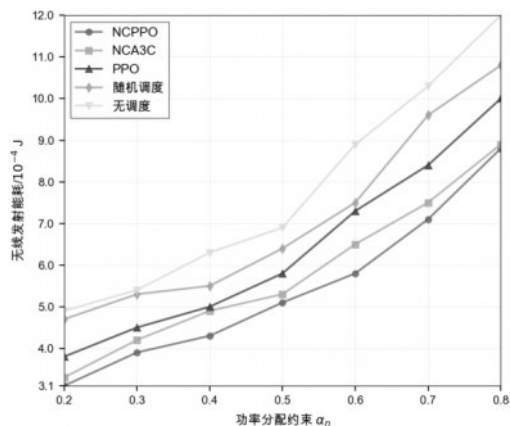


图12 不同算法在不同功率分配约束下的无线发射能耗对比

制不必要的能耗增长。上述结果表明,其能够控制功率资源充足时的发射开销。

## 5 结束语

本文面向Cell-Free上行连接优化,构建融合顺序门控与有限码长可靠性的随机服务模型,推导统计时延界,并以超时惩罚将其引入NCPPO。仿真表明,该方法在不同时延界、用户规模、功率约束及网络架构下均能降低时延超出量和发射能耗。后续将研究分布式实现、动态拓扑适应及工业互联网中的应用。



陈思曼 (2001—), 女, 北京邮电大学研究生, 主要研究方向为确定性时延、网络演算。



丰雷 (1987—), 男, 博士, 北京邮电大学副教授, 主要研究方向为去中心化边缘网络资源管理、工业蜂窝网络技术演进及应用。



杨志祥 (1996—), 男, 博士, 北京邮电大学博士后, 主要研究方向为蜂窝无源物联网。



汪旭东 (1999—), 男, 博士, 北京邮电大学博士后, 主要研究方向为生成无线网络优化、新型多址技术。



李文璟 (1973—), 女, 博士, 北京邮电大学教授, 主要研究方向为网络智能化运营管理、无线网络管理与优化。

## 参考文献:

- [1] Haque M E, Wang X, Liu Y. A Comprehensive Survey of 5G URLLC and Challenges in the 6G Era[EB/OL]. arXiv preprint arXiv: 2508.20205, 2025.
- [2] Zhu J, Wang S W. QoS-guaranteed resource allocation in mobile communications: a stochastic network calculus approach[J]. IEEE/ACM Transactions on Networking, 2024, 32(6): 5159-5171.
- [3] Zhang X, Wang J Q, Poor H V. Statistical delay and error-rate bounded QoS provisioning for mURLLC over 6G CF M-MIMO mobile networks in the finite blocklength regime[J]. IEEE Journal on Selected Areas in Communications, 2020, 39(3): 652-667.
- [4] Zhong Y, Zhou X, Feng K. Spatial network calculus: toward deterministic wireless networking[J]. IEEE Transactions on Wireless Communications, 2026, 25: 4707-4720.
- [5] Xu C, Liu Y, Wang C, Nallanathan A. Analysis of grant-free access for URLLC service: a stochastic network calculus perspective[J]. IEEE Transactions on Vehicular Technology, 2025, 74(9): 14949-14954.
- [6] Li D, Li J, Niyato D, Feng W, Jiang W. Deep energy-efficient optimization network for URLLC over cell-free massive MIMO[J]. IEEE Internet of Things Journal, 2025, 12(12): 20973-20987.
- [7] Chong B, Lu H, Guo F, Xue Z. Performance analysis for URLLC in cell-free massive MIMO systems with limited fronthaul capacity and hardware impairments[J]. IEEE Transactions on Wireless Communications, 2026, 25: 163-176.
- [8] Björnson E, Sanguinetti L. Making cell-free massive MIMO competitive with MMSE processing and centralized implementation[J]. IEEE Transactions on Wireless Communications, 2019, 19(1): 77-90.
- [9] Lee Y, Lee D, Oh J, Song C, Noh W, Cho S. Researches on resource allocation with reinforcement learning in cell-free networks[C]//2025 International Conference on Artificial Intelligence in Information and Communication (ICAIC). Fukuoka, Japan, 2025: 0151-0153.
- [10] Sun Y, Jiang X, Pan G, Zhang S, Xu S. Joint computation offloading and multi-user scheduling for max-latency minimization in MEC-aided cell-free networks[C]//2024 16th International Conference on Wireless Communications and Signal Processing (WCSP). Hefei, China, 2024: 703-708.
- [11] Thai T V, Nguyen H V, Le M T P, Tran H Q. Latency optimization in clustering NOMA-aided cell-free massive MIMO with mobile edge computing[C]//2025 8th International Conference on Circuits, Systems and Simulation (ICCSS). Ho Chi Minh City, Vietnam, 2025: 83-89.
- [12] Tilahun F D, Abebe A T, Kang C G. DRL-based distributed resource allocation for edge computing in cell-free massive MIMO network[C]//GLOBECOM 2022 - 2022 IEEE Global Communications Conference. Rio de Janeiro, Brazil, 2022: 3845-3850.
- [13] Li N N, Zhang Y, Wang H. Min - max latency optimization for IRS-aided Cell-Free mobile edge computing systems[J]. IEEE Internet of Things Journal, 2023, 11(5): 8757-8770.
- [14] Mahmoudi A, Zaher M, Björnson E. Joint energy and latency optimization in federated learning over cell-free massive MIMO networks[C]//2024 IEEE Wireless Communications and Networking Conference (WCNC). Dubai, United Arab Emirates, 2024: 1-6.
- [15] Siddiqui M U A, Hussain S, Ali A. URLLC in beyond 5G and 6G networks: An interference management perspective[J]. IEEE Access, 2023, 11: 54639-54663.
- [16] Fan Q, Li Y, Zhang J. Delay-aware resource allocation in fog-assisted IoT networks through reinforcement learning[J]. IEEE Internet of Things Journal, 2021, 9(7): 5189-5199.
- [17] Chong B, Lu H. Statistical delay performance analysis for URLLC in uplink cell-free massive MIMO systems: a stochastic network calculus perspective[J]. IEEE Transactions on Wireless Communications, 2025, 24(2): 1162-1177.
- [18] Li Z, Wang Y, Liu H. Stochastic Network Calculus-Aided Delay Analysis of Wireless-Power Line Mixed Networks[J]. Electronics, 2022, 11(9): 1326.
- [19] Cui P, Han S, Xu X. End-to-End Delay Performance Analysis of Industrial Internet of Things: A Stochastic Network Calculus Perspective[J]. IEEE Internet of Things Journal, 2024, 11(3): 5374-5387.
- [20] 王华华, 黄烨霞, 李玲. 基于 FL 的无蜂窝网络用户调度与功率分配策略[J]. 通信学报, 2024, 45(9): 129-143.  
Wang H H, Huang Y X, Li L. User scheduling and power allocation strategy for cell-free networks based on FL[J]. Journal on Communications, 2024, 45(9): 129-143.
- [21] Chen Y, Liu Y, Zhang H. Statistical QoS provisioning analysis and performance optimization in xURLLC-enabled massive MU-MIMO networks: A stochastic network calculus perspective[J]. IEEE Transactions on Wireless Communications, 2024, 23(7): 8044-8058.
- [22] Lancho A, Durisi G, Sanguinetti L. Cell-free massive MIMO for URLLC: A finite-blocklength analysis[J]. IEEE Transactions on Wireless Communications, 2023, 22(12): 8723-8735.
- [23] Huang Y, Jiang Y, Zheng F C, Zhu P, Quek T Q S. Effective energy efficiency of cell-free mMIMO systems for URLLC with probabilistic

- delay bounds and finite blocklength communications[J]. IEEE Transactions on Wireless Communications, 2025, 24(3): 2279-2296.
- [24] Polyanskiy Y, Poor H V, Verdú S. Channel coding rate in the finite blocklength regime[J]. IEEE Transactions on Information Theory, 2010, 56(5): 2307-2359.
- [25] Fidler M. An end-to-end probabilistic network calculus with moment generating functions[C]//2006 14th IEEE International Workshop on Quality of Service. 2006: 261-270.
- [26] Wang K L, Liu Y, Zhang H. Task-oriented delay-aware multi-tier computing in cell-free massive MIMO systems[J]. IEEE Journal on Selected Areas in Communications, 2023, 41(7): 2000-2012.
- [27] Tang Y, Agrawal S. Discretizing continuous action space for on-policy optimization[EB/OL]. arXiv:1901.10500v4.
- [28] Lowe R, Wu Y, Tamar A, et al. Multi-agent actor-critic for mixed cooperative-competitive environments[C]//Advances in Neural Information Processing Systems 30. 2017: 6379-6390.
- [29] Schroeder De Witt C, Gupta T, Makoviichuk D, et al. Is independent learning all you need in the StarCraft multi-agent challenge?[EB/OL]. arXiv:2011.09533, 2020.
- [30] Schulman J, Moritz P, Levine S. High-dimensional continuous control using generalized advantage estimation[EB/OL]. arXiv preprint arXiv:1506.02438, 2015.
- [31] Schulman J, Wolski F, Dhariwal P, et al. Proximal policy optimization algorithms[EB/OL]. arXiv:1707.06347, 2017.
- [32] Lin Y M, Zhang H, Liu Y. Joint Model Partitioning and Bandwidth Allocation for UAV-Assisted Space - Air - Ground - Sea Integrated Network: A Hybrid A3C-PPO Approach[J]. Entropy, 2026, 28(3): 337.

#### [作者简介]



周凡钦（1988—），男，博士，北京邮电大学副教授，主要研究方向为网络智能管理。