面向扩展现实应用的多用户多输入多输出上行链路及时吞吐量最大化

Ravi Sharan B A G、K. Pavan Srinath、Alvaro Valcarce Rial 和 Baltasar-Beferull Lozano

arXiv:2602.05751v1 [cs.IT] 5 Feb 2026

摘要

本文研究通过上行链路多用户多输入多输出（多用户多输入多输出）调度，为扩展现实（扩展现实）应用实现跨层及时吞吐量最大化。我们采用信息峰值年龄（峰值信息年龄）度量来刻画及时调度机会，并将其作为约束纳入网络侧优化问题，以反映用户满意度。由于该问题属于NP难问题，我们采用一种无信令、基于加权比例公平的迭代启发式算法，其中权重根据峰值信息年龄度量确定。大量数值仿真结果表明，在不牺牲整体系统吞吐量的情况下，所提算法在扩展现实（扩展现实）容量方面始终优于现有基线方法。

索引词

6G通信、比例公平调度、信息峰值年龄、扩展现实、多用户多输入多输出

预印本声明：本文已提交至 IEEE，拟供其考虑发表。版权可能在不另行通知的情况下转让，届时此版本可能不再可供查阅。

一、引言

涵盖混合（增强）现实（MR）和虚拟现实（VR）技术的扩展现实（XR）服务，正在塑造沉浸式多媒体应用的未来。要实现无缝的扩展现实体验，

Ravi Sharan 任职于德国的诺基亚贝尔实验室。K. P. Srinath 和 Alvaro V. Rial 任职于法国的诺基亚贝尔实验室。电子邮箱：{ravi.bhagavathula, pavan.koteshwar srinath, alvaro.valcarce rial}@nokia-bell-labs.com。Baltasar B. Lozano 任职于挪威的 SimulaMet 和阿格德尔大学。电子邮箱：baltasar@simula.no。Ravi Sharan 在挪威阿格德尔大学攻读博士期间，参与了本研究的部分工作。

无线蜂窝网络需要提供高速率且可靠的数据速率（≈ 30 – 150 Mb/s），同时保持低时延（≈ 5–30 ms）。在共址多用户 MR 场景中，用户会同时与彼此及其虚拟环境交互，因此这些要求进一步提高。无论如何，MR 应用对上行链路（UL）的要求与对下行链路（DL）的要求同样高。这是因为用户设备（UE）会持续向无线接入网（RAN）传输用于 3D 渲染的原始视频 [1]。因此，优化上行链路对于充分发挥这些扩展现实服务的潜力至关重要。

为了在上行链路中提供有保障的服务质量（QoS），第三代合作伙伴计划（3GPP）研究目前正在考虑采用包含时延特定信息的用户设备调度 [2]。体验质量（QoE）则通过提取视频帧级的应用感知信息，并将其作为分组级元数据嵌入媒体访问控制层（MAC）中单独处理 [3]。另一方面，随着物理层（PHY）相关方面的最新进展，多用户多输入多输出（MU-MIMO）正受到广泛关注，并被认真研究用于第六代（6G）网络 [4]。由于小区内干扰会影响物理层的预编码，进而影响体验质量（Qo(S)E），因此，MU-MIMO 中的用户设备调度极具挑战性。因此，对于扩展现实应用中的 MU-MIMO 上行用户设备调度，必须审慎考虑物理层-媒体访问控制层相关方面。

基于广为人知的比例公平（PF）指标的调度方案已广泛应用于商用第五代移动通信（5G）蜂窝网络 [5]。PF 指标旨在根据信道条件最大化系统吞吐量，同时维持用户设备之间的长期公平性。展望未来，要在扩展现实场景中成功应用 PF 指标，有必要对其进行与体验质量（Qo(S)E）相关的扩展。然而，获取上述应用感知信息会带来额外的上行链路信令开销，并且无法在不同应用之间实现完全通用化。在本文中，我们采用无信令方法，并将信息年龄（*AoI*）与 *PF* 指标结合，用于扩展现实场景中的多用户多输入多输出（*MU-MIMO*）上行用户设备调度。信息年龄定义为自用户设备上次生成的数据包成功传输以来所经过的时间，它同时反映了数据包交付间隔和数据包时延 [6]。此外，我们采用信息峰值年龄（PAoI）指标来建模用户设备的体验质量（Qo(S)E），因为它可以方便地与分组时延预算（PDB）结合 [7]。据我们所知，目前尚无相关工作研究在考虑完整多用户多输入多输出（MU-MIMO）物理层相关方面的情况下，针对 XR 流量进行信息年龄感知上行链路调度。

相关文献 **ture:** 当前关于扩展现实用户设备调度的文献主要聚焦于下行链路单用户 -

多输入多输出（单用户多输入多输出）场景。例如， [8] 提出了一种面向扩展现实和5G增强型移动宽带流量的服务质量感知型调度启发式算法。该算法根据近期引入的协议数据单元（PDU）集概念对调度指标进行建模。 [9]中的作者提出了两种调度启发式算法，通过将多个视频帧合并为一个实体，最大化满足需求的扩展现实用户设备数量。 [10] 提出了一种信令辅助方案，以满足目标体验质量和扩展现实关键绩效指标，其中gNodeB（下一代基站）会从用户设备收集周期性测量数据。上述方案利用数据包级元数据对下行调度指标进行建模；若要将其用于上行用户设备调度，可能难以避免引入额外的信令开销。首先，在扩展现实应用严格的时延要求下，信令开销可能会对可实现的体验质量造成不利影响。即使信令得到妥善处理，小区内干扰仍会给多用户多输入多输出中的用户设备调度带来额外挑战。

另一方面， [11] 和 [12]探讨了单天线用户设备短分组通信中的信息年龄感知上行链路调度。 [11] 提出了一种深度强化学习（DRL）方案，在单用户多输入多输出（SU-MIMO）场景下，以全网年龄信息总和约束为条件实现吞吐量最大化。最后， [12]提出了一种干扰近似Whittle指数策略，以即时期限为条件最小化全网年龄信息总和。然而，这些研究并未揭示XR流量下的信息年龄特性。在此背景下，我们接下来介绍系统模型，随后阐述问题构建、所提方法和数值结果。

II. 系统模型

我们考虑一个多用户多输入多输出场景，其中单个下一代基站在上行链路中为 N 扩展现实用户设备提供服务。令 $\bar{\mathrm{N}} := \{1, 2, \ldots, N\}$，$|\mathrm{N}| = N$ 表示下一代基站所服务的所有用户设备的集合。集合 N 中的部分用户设备会在同一时间-频率二维资源栅格中，于单个传输机会期间被复用。在时域（TD）中，传输机会以传输时间间隔（TTI）$t \in \mathrm{T} \subseteq \mathbb{N}^{+}$表示。一个 TTI 包含 N_{sym} 个正交频分复用（OFDM）符号。在频域中，有 F 个资源块（RB），每个 RB $f \in \mathrm{F} := \{1, 2, ..., F\}$跨越 N_{sc} 个子载波。因此，对于每个 TTI t，每个 RB f总共有 $I_f = N_{sc} \cdot N_{sym}$ 个资源元素（RE）。其中，RE 是二维资源栅格的最小单位。此外，我们假设下一代基站配备 N_G 条收发链（TRX），且在不失一般性的情况下，用户设备 $n, \forall n \in \mathrm{N}$ 配备 N_U 条 TRX。在任意 TTI t，一个用户设备

n 使用 $\lambda_n(t) \leq N_U$ 条空间流传输数据。因此，共同调度的用户设备集合 $\mathrm{N}'(t) \subseteq \mathrm{N}$ 中的空间流总数为 $\Lambda(t) = \sum_{n\in\mathcal{N}'(t)} \lambda_n(t)$。此外，¯在任意 TTI t，集合 $\mathrm{N}'(t)$ 中的空间层总数不超过 $\Lambda(t) \leq \Lambda \leq N_G$ 。

至于无线信道，我们假设其为平坦块衰落信道，即它在单个传输时间间隔 t 内保持恒定，并在不同传输时间间隔之间以相关方式变化。类似地，假设信道在单个资源块 f 上保持恒定，并在不同资源块之间变化。在传输时间间隔 t 以及资源元素 $i_f \in \mathrm{I}f := \{1, 2, \ldots, I_f\}$上，下一代基站接收到的复信号向量$\mathbf{y}(i_f; t) \in \mathbb{C}^{N_G\times 1}$ 由下式给出：

$$\mathbf{y}(i_f; t) := \sum_{n\in\mathcal{N}'(t)} \mathbf{H}_n(f; t)\mathbf{x}_n(i_f; t) + \boldsymbol{\eta}(i_f; t). \tag{1}$$

在上述表达式中，$\mathrm{H}_n(f; t) \in \mathbb{C}^{N_G\times L_n(t)}$是用户设备 u_n与下一代基站之间的有效上行链路信道。假设有效信道利用已知导频符号进行估计，并包含预编码操作。项$\mathbf{x}_n(i_f; t) \in \mathbb{C}^{L_n(t)\times 1}$是用户设备 u_n在资源元素$(i_f; t)$上传输的数据向量。最后，$\boldsymbol{\eta}(i_f; t) \in \mathbb{C}^{N_G\times 1}$统称为小区间干扰（如有）加上均值为000、协方差矩阵为$\mathrm{R}_{cov} \in \mathbb{C}^{N_G\times N_G}$的热噪声。请注意，有效信道矩阵的索引是相对于 f的，而其余各项则不同，这是上述信道假设的结果。用户设备 n 在资源元素 i_f上的均衡后有效信干噪比（信干噪比）估计记为 $\rho_n(i_f; t)$，它是$\mathrm{H}_n(f; t)$、$\mathrm{H}_{n'}(f; t)$、$\forall n' \in \mathrm{N}'(t) \setminus \{n\}$和$\mathrm{R}_{cov}$的函数（由于篇幅限制，此处省略表达式，见 [13, 式 (5)]）。用户设备 n的可达吞吐量记为 $Q_n(t) \in \mathbb{R}_{\geq 0}$，则可近似 [13] 如下：

$$Q_n(t) \leq \lambda_n(t) \cdot \left(\sum_{f\in\mathcal{F}} \sum_{i_f\in\mathcal{I}_f} \log_2\left(1 + \hat{\rho}_n(i_f; t)\right) \right), \tag{2}$$

其中，$\hat{\rho}_n(i_f; t)$ 表示按发射功率缩放的信干噪比。

下一代基站执行以下操作序列。在一个传输时间间隔 t开始时，所有¯相关用户设备都会向下一代基站发送缓冲区状态报告（BSR）$b_n(t) \in \mathrm{B} := [0, B_n]$，指示¯待传输的比特数。此处，$B_n$ 表示数据包允许的最大大小。本文使用“数据包”一词指代 3GPP 扩展现实规范中的数据突发，它由多个协议数据单元组成，这些单元根据底层视频帧分组，并以周期性方式到达。随后，下一代基站选择一组用户设备，并针对每个 $n \in \mathrm{N}'(t)$，使用二元变量 $\beta_n(t) = 1$表示调度决策。下一代基站还会分配

空间流数量 $\lambda_n(t)$以及共同调度的用户设备在 $\mathrm{N}'(t)$ 中使用的调制与编码方案（MCS）值。随后，共同调度的用户设备执行资源块选择和上行功率控制，并通过传输块（TB）发送部分数据包。用户设备的传输块比特大小与其分配的调制与编码方案值成正比。下一代基站会发送即时反馈$s_n(t) \in \{0, 1\}$，表示相应传输块接收成功或失败。请注意，本文不考虑重传。用户设备根据 $s_n(t)$值更新其缓冲区状态报告，并随时间重复这一过程。

本文考虑如下数据包生成与信息年龄建模。仅当当前数据包满足以下条件时，才会在用户设备 n 处生成一个长度为 B_n，使得 $B_n \leq B_n$ 的新数据包：

（a）已被下一代基站成功接收，或者（b）超过数据包时延预算 D，从而导致数据包失败。

此外，我们假设 $B_n > Q_n(t)$，$\forall n \in \mathrm{N}$，$\forall t \in \mathrm{T}$，这意味着用户设备需要若干个传输时间间隔才能成功传输其数据包。用户设备的信息年龄值记为 $\Delta_n(t) \in \mathbb{N}^+$，随时间线性增长，并且仅当该用户设备被调度进行传输且当前数据包成功传输时，才重置为其默认值（本文不失一般性地设为1）。我们还考虑对超过 Δ 的信息年龄进行有限值截断，且 Δ被选为大于 N 和D。下一代基站为每个用户设备维护一个独立的计数器，以跟踪信息年龄的演化，其可形式化表示为：

$$\Delta_n(t+1) := \min(\Delta_n(t), \bar{\Delta}) \cdot (1 - \phi_n(t+1)) + 1. \tag{3}$$

这里，$\phi_n(t+1)$ 是一个跨越当前和下一个传输时间间隔 $t \text{ and } t+1$ and 的联合伯努利随机变量，用于建模用户设备选择、传输块级即时反馈以及用户设备的数据包补充（下文将其表示为关于缓冲区状态报告的指示函数）。形式上，其定义如下：

$$\phi_n(t+1) := \begin{cases} 1, \text{ if } \beta_n(t) \wedge s_n(t) \wedge \mathbb{1}_{\{b_n(t+1) > b_n(t)\}} = 1 \\ 0, \text{ otherwise.} \end{cases} \tag{4}$$

式 (4) 中联合条件的重要意义在于，下一代基站可以根据上述数据包生成模型，推断用户设备的数据包补充是由条件 (a) 还是 (b) 导致的。这也免去了在扩展现实应用中进行上行用户设备调度时，用户设备与下一代基站之间交换除缓冲区状态报告以外的显式分组级元数据的需要。

除了信息年龄建模，我们还将时间平均信息峰值年龄作为一种指标，以有效评估体验质量和 XR 关键绩效指标。时间平均信息峰值年龄记为 $\Delta_{n,PAoI}$，是指在 $\phi_n(t) = 1$ 之前的信息年龄值的平均值。形式上，$\Delta_{n,PAoI}$ 可表示为：

$$\Delta_{n,PAoI} := \frac{\sum_{\mathcal{T}} \Delta_n(t)}{\sum_{\mathcal{T}} \phi_n(t+1)} \tag{5}$$

基于上述假设和建模，我们现在提出本文所研究的优化问题。

III. 及时吞吐量最大化

最大化网络中得到满足的用户设备数量（称为 XR 容量）是一项重要的 XR关键绩效指标。若用户设备至少有 99% 的数据包在数据包时延预算 [1] 内成功接收，则该用户设备被视为得到满足。对于第 II 节所述的系统模型，可以通过公平的用户设备调度来最大化系统吞吐量，从而部分实现上述关键绩效指标。此外，还可以利用式 (5) 中定义的峰值信息年龄度量，通过及时调度机会纳入用户设备满意度。本文将以此方式实现的吞吐量简称为及时吞吐量，并通过以下优化问题对其进行形式化表达：

$$\mathbf{P1:} \underset{\boldsymbol{\beta}(t);\forall t}{\text{maximize}} \ \limsup_{T\to\infty} \ \frac{1}{T} \sum_{\mathcal{T}} \sum_{\mathcal{N}} \beta_n(t) \cdot U_\alpha(Q_n(t)) \tag{6a}$$

$$\text{subject to} \left\{ \limsup_{T\to\infty} \ \Delta_{n,PAoI} \le \bar{D} \right\}_{n=1}^{\bar{N}}, \tag{6b}$$

$$\sum_{n\in\mathcal{N}} \lambda_n(t) \cdot \beta_n(t) \le \bar{\Lambda}, \tag{6c}$$

$$\boldsymbol{\beta}(t) \in \{0,1\}^{\bar{N}}, \forall n \in \mathcal{N}.$$

式 (6) 中提出的问题是一个整数非线性规划问题，涉及式 (6a) 和 (6b) 中的时间平均表达式，以及式 (6c) 中的瞬时约束。优化变量 $\boldsymbol{\beta}(t) \in \{0, 1\}^N$ 是一个二进制向量，表示在传输时间间隔 t 内共同调度的用户设备。式 (6a) 中的函数$U_\alpha(.)$ 表示众所周知的 α-公平效用函数 [14]，用于确保用户设备调度的公平性。用户设备的信息峰值年龄 $\Delta_{n,\ PAoI}, \ \forall n \in$ N 与数据包时延预算 $\bar{D}$ 之间的相互关系由式 (6b) 中的约束集刻画。信息峰值年龄与数据包时延预算之所以能够耦合，是因为用户设备的信息年龄重置为默认值的时刻意味着数据包补充。此外，与单一

面向全网的用户设备信息峰值年龄效用函数相比，我们采用多个逐用户设备约束，以更好地反映 XR 关键绩效指标。最后，耦合约束 (6c) 确保下一代基站服务的空间层总数不超过 Λ。

可以观察到，单独最大化 (6a) 会将 **P1**简化为仅针对物理层方面的经典公平性保障多用户多输入多输出吞吐量最大化问题。然而，纳入 (6b) 后，**P1** 转变为一个跨层优化问题，可在数据包时延预算内为用户设备提供及时且有意义的数据包吞吐量。从求解角度来看，**P1**似乎可以通过放宽 (6c) 中的耦合约束，在用户设备之间解耦。然而，由于小区内干扰通过 $Q_n(t), \forall n \in \mathrm{N}$ 计算隐式地体现为耦合约束，解耦 (6) 仍非易事。此外，优化变量 $\boldsymbol{\beta}(t)$ 的整数属性使上述问题成为 NP 难问题。为有效应对这些挑战，我们提出一种具有实际应用价值的、按传输时间间隔执行的迭代启发式算法来求解 (6)，并使用信息峰值年龄加权优先级调度度量评估每次迭代的决策。

IV. 所提调度启发式算法

从第 III 节可以明显看出，求解（6）的用户设备调度方案应在用户设备公平性、系统吞吐量和满足信息峰值年龄约束之间进行有效权衡。广泛采用的基于比例公平度量的方案旨在不同信道条件和干扰条件下，在前两项指标之间取得良好平衡。对于（6），采用 PF度量对应于 $\alpha = 1$ in $U_\alpha(.)$ 的情形，并将（6a）转化为在每个传输时间间隔内最大化全网各用户设备吞吐量对数效用之和，即 $\sum_{n\in\mathcal{N}} \log(Q_n(t))$。虽然设置 $\alpha > 1$ 可以进一步提升公平性，但这可能会妨碍向用户设备提供及时的调度机会，并影响可实现的体验质量和 XR 关键绩效指标。这是因为，平均而言，信道条件较差的用户设备最终会在时域中获得很高比例的调度机会，因为 $\alpha \to \infty$。此外，这还可能导致小区内干扰，并最终因满足第 II 节数据包生成模型的条件（ b ）而使用户设备补充数据包。综合考虑以上因素，我们采用基于加权比例公平度量的调度启发式算法来求解（6），其中权重根据每个用户设备的信息峰值年龄度量确定。

为通过设置 $\alpha = 1 \text{in } U_\alpha(。)$，按传输时间间隔最大化 (6a)，即 $\sum_{n\in\mathcal{N}} \log(Q_n(t))$，需要最大化 $\sum_{n\in\mathcal{N}} Q_n(t) Q_{n,\,Avg}(t)$[14]。其中， $Q_n(t)$ 是式 (2) 中传输时间间隔 t 时每个用户设备可实现的吞吐量，该吞吐量取决于共同调度的用户设备子集 $\mathrm{N}'(t)$。此外，$Q_{n,\,Avg}(t)$ 是截至传输时间间隔 t 的每个用户设备已实现吞吐量的线性估计，其更新

规则如下：

$$Q_{n,Avg}(t+1) = (1-\tau)Q_{n,Avg}(t) + \tau \cdot \beta_n(t) \cdot Q_n(t). \tag{7}$$

此处，τ 是一个时间常数，其取值通常选为约1000个传输时间间隔的量级（例如$\tau = \frac{1}{1000}$）。我们现在将其扩展到(6)，并通过每个用户设备的乘法权重$W_{n,PAoI}(t)$对$Q_{n,Avg}(t)$进行缩放，从而将约束集(6b)纳入PF度量，定义如下：

$$W_{n,PAoI}(t) := \begin{cases} (\Delta_{n,wa}(t))^{-\kappa}, & \text{if } \Delta_{n,wa}(t) \le \bar{D} \quad (8a) \\ 1 - (\Delta_{n,wa}(t))^{-\kappa}, & \text{otherwise.} \quad (8b) \end{cases}$$

此处，$\Delta_{n,wa}$ 是相对于瞬时信息年龄 $\Delta_n(t)$以及峰值信息年龄度量的滑动平均$\Delta_{n,PAoI}$ 计算得到的加权平均，如下：

$$\Delta_{n,wa}(t) := \theta \cdot \Delta_n(t) + (1-\theta) \cdot \sum_{t'=0}^{t} \Delta_{n,PAoI}(t'). \tag{9}$$

量 $\kappa \in \mathbb{R}^+$ 与 $\theta \in [0,1]$ 是相互依赖的可调参数。例如，若在(9)中优先考虑$\Delta_{n,wa}(t)$，则 κ 取较小值；若优先考虑$\sum_{t'=0}^{t} \Delta_{n,PAoI}(t')$，则 κ取较大值。最后，(8)和(9)分别刻画了(6b)的右端和左端。

除了对加权信息年龄的PF度量进行建模外，我们还利用 Λ 与 $\mathbf{\Lambda}$之间的相互关系，在所提方法¯调度启发式算法中纳入了（6c）。更具体地说，我们发现，（6c）本质上转化为可在一个传输时间间隔 N（t） $\le N$内共同调度的用户设备数量上界¯，可近似表示如下：

$$N(t) := \left\lceil \frac{\bar{\Lambda}}{\frac{1}{\bar{N}} \sum_{n \in \mathcal{N}} \lambda_n(t)} \right\rceil, \implies \mathbf{1}^T \boldsymbol{\beta}(t) \le N(t), \tag{10}$$

所提调度启发式算法的伪代码见算法1，其中术语 $Q_{sum}(\mathcal{N}'(t)) := \sum_{n \in \mathcal{N}'(t)} Q_n(t)\, Q_{n,Avg}(t) \cdot W_{n,PAoI}(t)$。所提调度启发式算法在每个传输时间间隔为 t 选择使 Q_{sum} 最大的用户设备子集；也就是说，它优先调度 $Q_n(t)$ 值较高且乘积 $Q_{n,Avg} \cdot W_{n,PAoI}$ 值较低的用户设备。因此，式（8）中的 $W_{n,PAoI}$¯ 会引导 PF度量优先调度某个用户设备，直到 $\Delta_{n,wa}(t) \le D$，之后该度量将线性趋近于经典 PF度量。因此，所提调度启发式算法在满足信息峰值年龄约束、系统吞吐量和用户设备公平性之间实现了良好平衡。

算法1 每TTI及时吞吐量最大化 初始化： $\tau, \kappa, \theta, \Lambda, D$。输入： $\Delta_{n,wa}(t)$, $\mathrm{H}_n(f;t)$, $\mathrm{R}(i_f;t)$, $\forall i_f \in \mathrm{I}_f, \forall f \in \mathrm{F}$, $\forall n \in \mathrm{N}$。1: 设置 $\mathrm{N}'(t) \leftarrow \varnothing$, $\mathrm{K}(t) \leftarrow \varnothing$, 以及 $Q_{sum}(\mathrm{N}'(t)) \leftarrow 0$。2: 对于每个 $n \in \mathrm{N}$, 执行 3: $\mathrm{K}(t) \leftarrow \mathrm{N}'(t)$ 4: 如果 $Q_{sum}(\mathrm{N}'(t) \cup \{n\}) > Q_{sum}(\mathrm{N}'(t))$, 则 5: $\mathrm{N}'(t) \leftarrow \mathrm{N}'(t) \cup \{n\}$ 6: $Q_{sum}(\mathrm{N}'(t)) \leftarrow Q_{sum}(\mathrm{N}'(t) \cup \{n\})$ 7: 结束如果 8: 如果 $|\mathrm{N}'(t)| > N(t)$ 或 $\sum_{n \in \mathrm{N}'(t)} \lambda_n(t) > \Lambda$, 则 9: 返回 $\mathrm{K}(t)$。

▷ 提前停止规则。

10: **否则**继续 11: **结束 if**12: **结束循环**
13: **返回** $\mathrm{N}_\prime(t)$.

▷ 等价于在式（6）中求得 $\boldsymbol{\beta}(t)$。

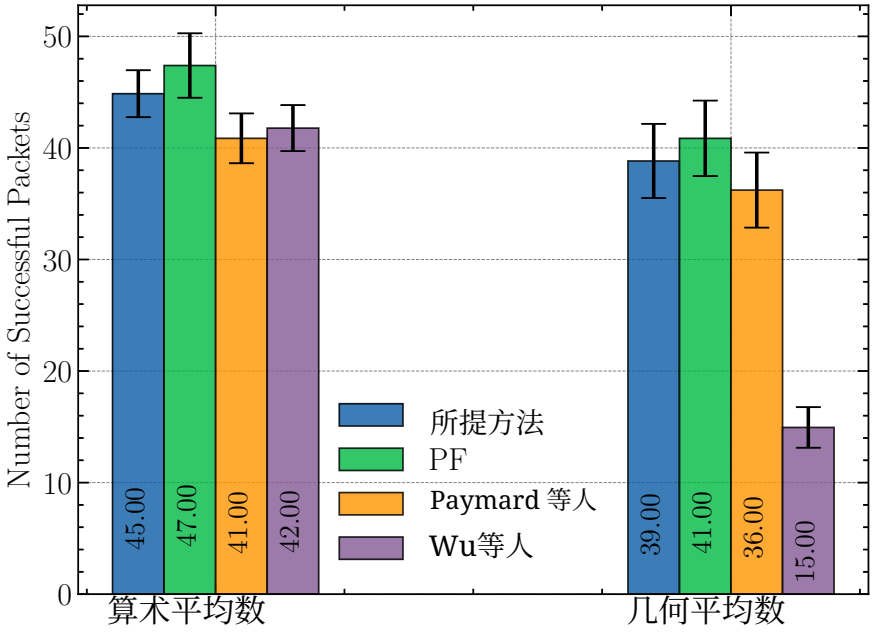


图1：排名前 95% 用户设备的有效吞吐量对比

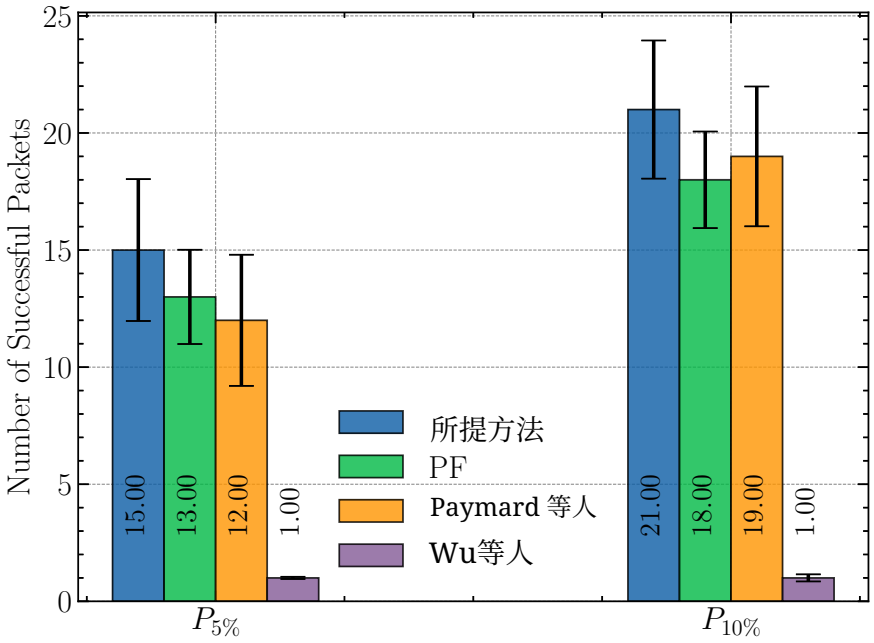


图 2：底部 5% 和 10% 用户设备的有效吞吐量比较

V. 数值结果

在本节中，我们使用以 NumPy（用于信道估计）和 TensorFlow 编写的多小区（本文中为下一代基站）、多链路级仿真器，评估算法1的性能。¯我们考虑一个包含三个下一代基站的单站点，其中每个下一代基站均配置为以多用户多输入多输出模式服务 $N = 10$ 个用户设备。每个下一代基站使用 48 个（3×1 子阵列）天线单元，并配置为 N_G = 16-收发器（$2\text{H} \times 4\text{V} \times 2\text{P}$）。此外，我们考虑下一代基站在任一传输时间间隔可支持的最大空间¯层数，即 $\Lambda = 8$。用户设备配置为 N_U = 4-收发器（$2\text{H} \times 1\text{V} \times 2\text{P}$）。对于物理层属性，我们采用 3GPP 38.901 城市微蜂窝（UMi）非视距信道模型，并使用第五代移动通信导频进行信道估计。其他相关物理层属性，如用户设备发射功率、资源块设置、调制与编码方案表以及信道编码参数等，可见 [13,第 5 节，表 I]）。此外，我们还采用 [13, 第 5.A 节] 中的基线上行功率控制、资源块分配和用户设备秩选择。对于扩展现实应用属性，为简化评估，我们考虑数据包-¯大小为 $B_n(t) = B = 75000$ 比特（9375 字节），$\forall n \in$ N。数据包时延预算值¯设为 $D = 30$，并在数据包时延预算之外额外考虑两个传输时间间隔的宽限期。这是为了让用户设备能够在数据包补充之前发送其最后剩余的传输块（如果有）。对于缓冲区状态报告，我们遵循 [15, 表 6.1.3.1-1]。此外，信息新鲜度截断值设为

¯作为 $\Delta = 100$。式 (8) 和 (9) 中的可调参数取为 $\kappa = 2$ 和 $\theta = 0$。5。最后，仿真共进行 35 次单个丢弃，传输时间间隔总数为 10^6 。对于基线方案，我们考虑了 [11] 中的深度强化学习解决方案，并采用算法1中的经典 PF度量和加权 PF 度量。

我们首先在图1和图 2中评估有效吞吐量，即用户设备处数据包补充的平均数量，并比较其 90% 置信区间。这里，平均值同时针对传输时间间隔和用户设备数量计算，并四舍五入到最接近的整数。如图1所示，对于排名前 95% 的用户设备，所提调度启发式算法在 [8] 和 [11] 方面均优于基线方案，其算术平均数和几何平均数皆如此。同时，所提调度启发式算法相较于 PF 基线也取得了相当的性能。另一方面，如图 2 所示，对于排名后 5% 和 10% 的用户设备，或小区边缘用户设备，所提启发式算法始终优于所有基线方案。这表明，通过式 (8) 中的峰值信息年龄权重对 PF度量进行缩放，只会略微牺牲排名前 95% 用户设备的有效吞吐量，从而兼顾小区边缘用户设备。此外，这种方法效率很高，因为小区边缘用户设备使用较低的调制与编码方案值进行传输（可能造成干扰），并且平均而言需要更多调度机会才能成功传送数据包。

接下来，我们通过图 3 评估所提调度启发式算法和基线方法所实现的 XR容量。具体而言，我们使用箱线图比较用户设备时间平均信息峰值年龄的分布。箱线图能够简洁地概括分布，其中数据点要么按四分位数划分，要么在超出四分位数范围时标记为离群值 [16]。图 3 中的 q 表示满足约束 (6b) 的用户设备百分比。首先，所提调度启发式算法实现了最高的 XR 容量，为$q = 81\%$。与经典PF调度器相比，这是显著的提升，尤其是考虑到二者的有效吞吐量表现相当。此外，与 [8] 中的基线方法相比，XR 容量的合理提升仍然意义重大，因为所提调度启发式算法采用无信令方法，而 [8]则不然。其次，所提调度启发式算法实现了最低的 $\mu = 28$分布均值，远低于数据包时延预算。这是因为，与离群值尾部分布较长的基线方法相比，所提调度启发式算法对应的离群值更接近第四四分位数。此外，尽管经典PF调度器和 [8] 中的基线方法实现的均值接近数据包时延预算，但由于尾部分布较长，其 XR 容量仍未达到最大值。这表明，相较于全网PAoI总和最小化，式 (6b) 中的逐用户设备约束是必要的。最后，它可以

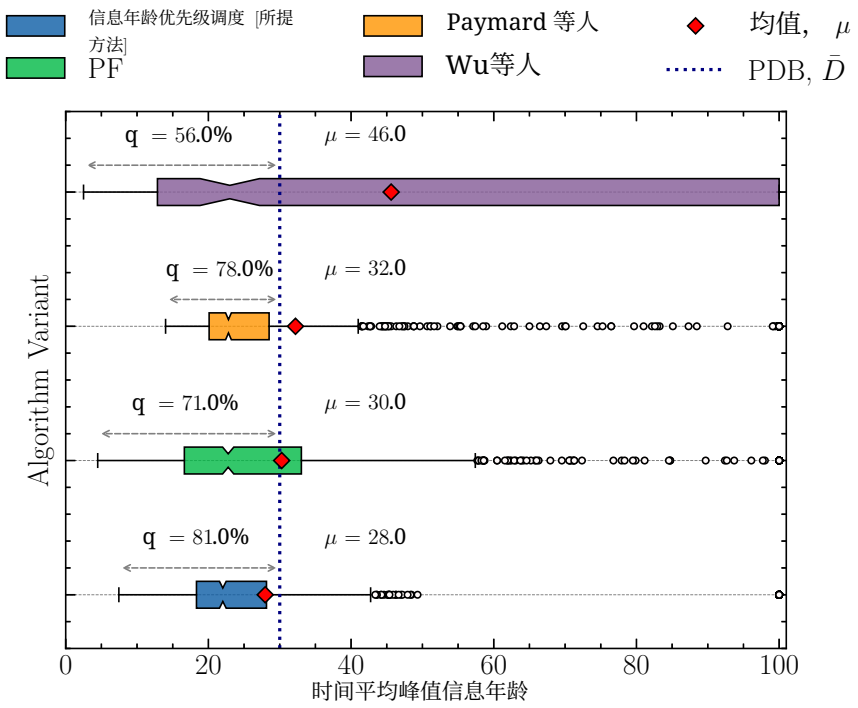


图 3：PAoI分布比较

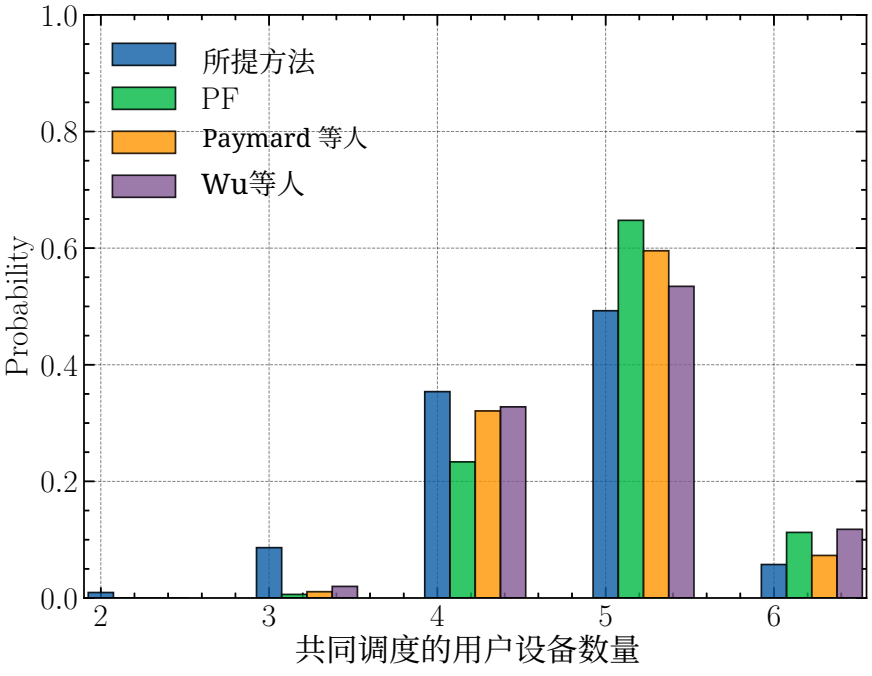


图4：共同调度的用户设备的经验分布

可以观察到，所提调度启发式算法在 [11] 方面的有效吞吐量和XR容量均优于深度强化学习基线。这是因为深度强化学习基线缺乏关于多用户多输入多输出所导致干扰的明确信息，且无法适应式（6）中的XR关键绩效指标。

在图4中，我们比较了各调度方案下共同调度的用户设备数量的概率质量函数（pmf）。可以观察到，式（8）中的信息峰值年龄权重会在必要时调节调度的用户设备数量。这与各基线方法形成对比：基线方法中共同调度的用户设备数量更多集中在较高取值上。

VI. 结论

在本研究中，我们提出了一种迭代式用户设备调度启发式算法，以解决具备XR关键绩效指标的上行链路多用户多输入多输出中的及时吞吐量最大化问题。数值仿真表明，所提启发式算法在不牺牲已实现吞吐量的情况下，能够实现最高的XR容量。我们认为，可以利用深度强化学习技术降低所提方法的计算复杂度。

参考文献

[1] Margarita Gapeyenko、Vitaly Petrov、Stefano Paris、Andrea Marcano 和 Klaus I Pedersen。《5G 和 5G-Advanced 3GPP 新空口中的扩展现实（XR）标准化》。*IEEE Network*，37(4):22–28，2023。[2] 3GPP。《扩展现实与媒体服务（XRM）第二阶段研究，Release 19 SA2》。技术报告，第三代合作伙伴计划（3GPP），2024年1月。[3] Abolfazl Amiri、Pouria Paymard、Pilar Andres-Maldonado、Stefano Paris、Klaus I. Pedersen 和 Troels Kolding。《扩展现实服务的应用感知：5G-Advanced 及其未来》。*IEEE Communications Magazine*，62(8):38–44，2024。[4] Jeffrey G. Andrews、Todd E. Humphreys 和 Tingfang Ji。《6G 初具雏形》。*IEEE BITS the Information Theory Magazine*，4(1):2–24，2024。[5] Hao Yin、Lyutianyang Zhang 和 Sumit Roy。《5G NR 中增强型移动宽带业务内的超可靠低时延通信流量复用：公平调度》。*IEEETransactionsonCommunications*，69(2):1080–1093，2021。[6] Roy D. Yates、Yin Sun、D. Richard Brown、Sanjit K. Kaul、Eytan Modiano 和 Sennur Ulukus。《信息年龄：导论与综述》。*IEEE Journal on Selected Areas in Communications*，39(5):1183–1210，2021。[7] Kumar Saurav 和 Rahul Vaze。《峰值信息年龄约束下的在线能量最小化》。载于*2021*年第四届移动、自组织和无线网络建模与优化国际研讨会（*WiOpt*），第1–8页，2021。[8] Pouria Paymard、Stefano Paris、Abolfazl Amiri、Troels E. Kolding、Fernando Sanchez Moya 和 Klaus I. Pedersen。《多业务5G-Advanced网络中面向XR流量的PDU集合调度算法》。载于*ICC 2024 - IEEE*国际通信会议，第758–763页，2024。[9] Erkai Chen、Shengyue Dou、Shaobo Wang、Youlong Cao 和 Shuri Liao。《5G及后续网络中的扩展现实帧级集成传输》。载于*2021*年*IEEE*全球通信大会（*GLOBECOM*），第1–6页，2021。

[10] Biljana Bojovi´c、Sandra Lag´en、Katerina Koutlia、Xiaodi Zhang、Ping Wang 和 Liwen Yu。《通过扩展现实环回机制提升面向XR流量的5G服务质量管理》。*IEEE Journal on SelectedAreasinCommunications*，41(6):1772–1786，2023。[11] Chien-Cheng Wu、Petar Popovski、Zheng-Hua Tan 和 ˇCedomir Stefanovi´c。《利用强化学习设计感知信息年龄的5G上行调度器》。载于*2021*年第四届*IEEE 5G*世界论坛（*5GWF*），第176–181页，2021。[12] Song Li、Min Li、Ruirui Chen 和 Yanjing Sun。《多天线蜂窝物联网状态更新的传输年龄最优调度》。《中国通信》，19(4):302–314，2022。[13] K Pavan Srinath、Alix Jeannerot 和 Alvaro Valcarce Rial。《采用线性收发器的多用户多输入多输出系统中的联合资源与功率分配及用户设备秩选择》。*arXiv* 预印本 *arXiv:2407.16483*，2024。[14] M. Uchida 和 J. Kurose。《加权α比例公平性的信息论刻画》。载于*IEEEINFOCOM2009*，第1053–1061页，2009。[15] 3GPP。《新空口；介质访问控制（MAC）协议规范》。技术规范（TS）38.321，第三代合作伙伴计划（3GPP），2025年4月。版本18.5.0。[16] Hadley Wickham 和 Lisa Stryjewski。《箱线图的40年》。技术报告，had.co.nz，2012。