第五代移动通信技术移动IAB节点的分层资源分配学习与多智能体协调

Mohamed Sana、Benoit Miscopein，CEA-Leti，
Universit´eGrenoble Alpes，F-38000 格勒诺布尔，法国 电子邮件：{mohamed.sana, benoit.miscopein}@cea.fr

arXiv:2302.07573v1 [cs.MA] 15 Feb 2023

摘要—我们研究一种采用集成接入与回传的动态毫米波网络，其中移动中继节点会移动，以自动重新配置无线回传。具体而言，我们关注带内中继 网络，其在同一频段上承载接入与回传链路，并受到严重的同信道干扰限制。在此背景下，我们联合研究动态中继节点定位、用户关联和回传容量分配这一复杂问题。为以有限的复杂度解决该问题，我们采用具有两级结构的分层多智能体强化学习。高层策略动态协调移动中继节点，为低层策略定义回传配置；低层策略则联合为每个中继分配用户设备，并相应地分配回传容量。由此得到的解决方案能够自动适应用户数量、流量分布和信道变化，从而调整接入与回传网络。数值结果表明，我们提出的解决方案在分层学习过程的收敛性方面效果显著。与基准方法相比，该方案还显著提高了回传容量和网络总速率（最高达 3.5×）。

I. 引言

对数据吞吐量要求较高（峰值高达 20 Gbps）的增强型移动宽带服务（eMBB），是近期标准化的5G网络的主要目标之一 [1]。为提升网络容量，5G采用蜂窝致密化与毫米波通信相结合的方式，以利用这些频段可用的丰富频谱资源 [1]。此外，在地理区域内进行频谱空间复用，可使蜂窝致密化显著提升毫米波基站的覆盖质量以及小区边缘用户的性能 [2]。然而，致密化给无线资源管理带来了严峻挑战；随着用户设备和基站数量增加，无线资源管理也会变得更加复杂。此外，由于同信道干扰和回传容量有限，网络容量不会随着蜂窝致密化而必然提升，因此需要相应地增加回传容量。然而，回传网络通常依赖有线光纤或微波链路，部署成本高昂，从而限制了回传容量，并可能影响网络频谱效率和终端用户的服务质量（QoS）。

为解决这一问题，第五代移动通信技术还引入了集成接入与回传（IAB）网络，作为有线回传网络的一种经济高效的替代方案 [3], [4]。事实上，毫米波频率下可用的频谱资源十分丰富，因此能够将总带宽划分为不同部分，分别用于无线接入网络和回传网络 [3]。在所考虑的系统模型中，多个移动中继站（以下简称mIAB节点）动态移动，形成无线回传网络，并与覆盖其上的毫米波基站（以下简称IAB供体）协同，为多个已部署的移动用户设备提供接入服务。本研究聚焦于带内中继IAB网络，这类网络在同一频段上同时进行接入与回传链路传输。由于接入与回传链路会相互干扰，此类网络面临严格的干扰约束。因此，需要对mIAB节点进行动态协调，并在接入网络和回传网络上实施高效的联合无线资源管理。这一问题已受到学术界和业界的广泛关注 [5]–[8]。在 [5], 中，作者提出了一种集中式算法，在考虑回传网络负载的基础上优化用户关联。 [9] 提出了一种采用Q 学习算法的类似方法。在 [10],中，作者提出了一种用于频谱分配的深度强化学习（RL）算法，研究对象为带外中继IAB网络。 [11]中的作者提出了一种基于多智能体强化学习（MARL）的算法，以应对用户设备的移动性。然而，这些研究均未考虑移动IAB节点。与之不同，我们研究mIAB节点的动态定位问题，以重新配置回传网络，同时进行资源分配，即接入网络上的用户关联和回传容量分配。此外，我们还考虑了干扰、用户设备的移动性、IAB节点的负载以及用户设备业务请求随时间变化等环境动态性，这进一步增加了寻找精确且易于处理的解决方案的难度。例如， [12]采用穷举搜索算法来优化采用无人机的mIAB网络，但该算法在实际应用中难以处理。

为以有限的复杂度解决这一问题，我们采用一种基于分层 强化学习（hRL）的方法 [13]。该层次结构通过将上述问题分解为两个子问题而形成。高层优化负责动态协调移动式集成接入与回传节点，并确定其位置，从而共同最大化回传容量和用户设备覆盖。随后，低层优化联合确定接入与回传网络的最优资源分配。我们在多智能体强化学习（MARL）框架内处理这两项优化，其中将每个mIAB节点和每个用户设备建模为独立智能体，仅依赖对无线环境的少量观测来管理无线资源。我们提出了专门用于以有限复杂度学习分层策略的新奖励函数。所提出的方案在设计上具有灵活性和可扩展性，并能根据相对于（w.r.t.）环境动态性以及接入网络的需求，自动重新配置回传网络。

本研究得到了法国政府复苏计划（CRIIOT 项目）以及 H2020 DEDICAT 6G 项目（编号：101016499）的支持。

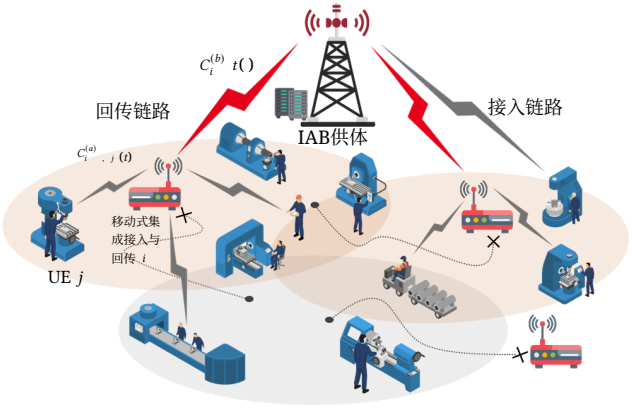

图1. 由1个IAB供体部署3个IAB节点的系统模型

II. System Model

我们考虑如图1所示的下行毫米波网络，该网络由一个固定的IAB供体组成，后者与 N_s 个mIAB节点共同构成回传网络；这些节点在网络中移动，为时间t的 $K(t)$个用户设备提供接入，服务对象为一组 $\mathcal{U}(t) = \{1,\dots,K(t)\}$。我们用 $\mathcal{S}_0 = \{1,\dots,N_s\}$表示mIAB节点集合，用 $\mathcal{S} = \mathcal{S}_0 \cup \{0\}$ 表示所有IAB站点的集合，其中包括索引为0的IAB供体。在这一动态网络中，每个配备接收天线的mIAB节点 $i \in \mathcal{S}_0$都可以在空间区域 $\mathcal{L} \subset \mathbb{R}^2$ 内调整其二维位置 $\ell_i(t)$，以与IAB供体建立无线回传链路。定位后，由于波束成形能力有限，每个mIAB节点 i 最多可同时为 L_i 个用户设备提供服务。我们分别用 L_0 和 M 表示IAB供体下行接入链路和回传链路的最大数量。最后，给定 $\rho_i(t)$，即IAB站点$i \in \mathcal{S}$ 在时间 t的负载（活跃用户设备数量），我们定义 $\kappa(t)$表示接入网络随时间变化的波束覆盖率（即总波束利用率）：

$$\kappa(t) = \frac{\sum_{i\in\mathcal{S}} \rho_i(t)}{\min\left(K(t), \sum_{i\in\mathcal{S}} L_i\right)}, \tag{1}$$

A. 接入与回传容量

由于我们关注的是带内 中继IAB网络，接入链路与回传链路在频率上部分重叠。我们假设回传网络和接入网络采用空分多址（SDMA）。在我们的系统模型中，IAB供体将所有可用的毫米波带宽 B 分配给其服务的mIAB节点，而这些节点则以 μ双工模式 [14],运行，所有接入链路仅使用频带的一部分 μB ，其中 $\mu \in [0,1]$。在此情形下，由于我们没有专门优化波束成形器，工作波束的重叠会造成干扰。因此，当用户设备 j 从IAB站 i接收数据时，会经历下行信号与干扰加噪声比（SINR）$\gamma_{i,j}^{(a)}(t)$，其表达式如下：

$$\mathrm{SINR}_{i,j}^{(a)}(t) = \frac{\chi_{i,j}(t) P_{i,j}^{\mathrm{Tx}} G_{i,j}^{\mathrm{Tx}}(t) G_{i,j}^{\mathrm{H}}(t) G_{i,j}^{\mathrm{Rx}}(t)}{I_{i,j}^{(a)}(t) + \mu N_0 B}. \tag{2}$$

这里，P_i^{Tx}，$_j$表示从 IAB 站 i到用户设备 j的发射功率， N_0 表示高斯噪声功率谱密度，而 G_i^{Tx}，$_j(t)$ 和 G_i^{Rx}，$_j(t)$ 分别表示IAB 站 i 与用户设备 j之间的发射天线增益和接收天线增益。此外，$\chi_{i,}$ $_j(t)$表示小尺度衰落系数，$G_{i,}^{\mathrm{H}}$ $_j(t)$表示信道增益，用于反映距离相关的路径损耗和大尺度阴影效应。特别要注意，$G_{i,}^{\mathrm{H}}$ $_j(t)$受 mIAB节点 i的位置 $\ell_i(t)$影响。最后，由于我们假设所有毫米波链路均采用全空间频率复用，与 IAB 站 i通信的用户设备 j 所经历的总干扰 $I_{i,}^{(a)}$ $_j(t)$，来自接入链路的小区内和小区间干扰，以及回传链路的小区间干扰[1]：

$$\begin{aligned} I_{i,j}^{(a)}(t) &= \chi_{i,j}(t) \sum_{j'\in\mathcal{U}\setminus\{j\}} x_{i,j'}(t) P_{i,j'}^{\mathrm{Tx}} G_{(i,j')\to j}^{\mathrm{Tx}}(t) G_{i,j}^{\mathrm{H}}(t) G_{i,j}^{\mathrm{Rx}}(t) \\ &+ \sum_{i'\in\mathcal{S}\setminus\{i\}} \sum_{j'\in\mathcal{U}\setminus\{j\}} x_{i',j'}(t) \chi_{i',j}(t) P_{i',j'}^{\mathrm{Tx}} G_{(i',j')\to j}^{\mathrm{Tx}} G_{i',j}^{\mathrm{H}}(t) G_{(i,j)\leftarrow i'}^{\mathrm{Rx}} \\ &+ \chi_{0,j}(t) \mu \sum_{i'\in\mathcal{S}_0} z_{i'}(t) P_{0,i'}^{\mathrm{Tx}} G_{(0,i')\to j}^{\mathrm{Tx}} G_{0,j}^{\mathrm{H}}(t) G_{(i,j)\leftarrow 0}^{\mathrm{Rx}}(t). \end{aligned} \tag{3}$$

其中，$x_{i,}$ $_j(t)$为二元用户关联变量；若用户设备 j 与集成接入与回传节点 i关联，则其值为1，否则为0。同样， $z_i(t)$为回传链路关联变量；若集成接入与回传节点 i 与IAB供体之间的回传链路处于活动状态，则其值为1，否则为0； $G_{(u,v)\to w}^{\mathrm{Tx}}$ 表示当 v 从u接收数据时， u 朝向 w的发射天线增益； $G_{(u,v)\leftarrow w}^{\mathrm{Rx}}$ 表示当 v 从 u 接收数据时， w 朝向 v的接收天线增益。因此，IAB站 i 与用户设备 j之间的接入链路容量记为 $C_{i,}^{(a)}$ $_j(t)$，其表达式为：

$$C_{i,j}^{(a)}(t) = \mu B \cdot \log_2\left(1 + x_{i,j}(t)\mathrm{SINR}_{i,j}^{(a)}(t)\right). \tag{4}$$

同理，mIAB节点 $i \in \mathcal{S}_0$ 与 IAB供体之间回传链路的总容量 $C_i^{(b)}(t)$ 表示为：

$$C_i^{(b)}(t) = B \cdot \log_2\left(1 + z_i(t)\mathrm{SINR}_i^{(b)}(t)\right), \tag{5}$$

其中，经历的信干噪比$\gamma_i^{(b)}(t)$ $\forall i \in \mathcal{S}_0$ 如下所示：

$$\mathrm{SINR}_i^{(b)}(t) = \frac{\chi_{0,i}(t) P_{0,i}^{\mathrm{Tx}} G_{0,i}^{\mathrm{Tx}}(t) G_{0,i}^{\mathrm{H}}(t) G_{0,i}^{\mathrm{Rx}}(t)}{I_i^{(b)}(t) + N_0 B}. \tag{6}$$

此处与式 (3) 不同， $I_i^{(b)}(t)$包含小区内回传链路干扰、接入链路干扰以及相应 mIAB节点同时收发所产生的自干扰[1]：

$$\begin{aligned} I_i^{(b)}(t) &= \chi_{0,i}(t) \sum_{i'\in\mathcal{S}\setminus\{i\}} z_{i'}(t) P_{0,i'}^{\mathrm{Tx}} G_{(0,i')\to i}^{\mathrm{Tx}}(t) G_{0,i}^{\mathrm{H}}(t) G_{0,i}^{\mathrm{Rx}}(t) \\ &+ \mu \sum_{i'\in\mathcal{S}\setminus\{i\}} \sum_{j'\in\mathcal{U}} x_{i',j'}(t) \chi_{i',i}(t) P_{i',j'}^{\mathrm{Tx}} G_{(i',j')\to i}^{\mathrm{Tx}} G_{i',i}^{\mathrm{H}}(t) G_{(0,i)\leftarrow i'}^{\mathrm{Rx}} \\ &+ \mu \sum_{j'\in\mathcal{U}} z_i(t) x_{i,j'}(t) P_{i,j'}^{\mathrm{Tx}} \xi_i, \end{aligned} \tag{7}$$

其中， ξ_i 是mIAB节点 i 的自干扰抵消增益，表示干扰抵消前后接收到的自干扰功率之比。

[1]接入与回传链路之间的相互干扰仅对重叠频带求和。相应的幅度因子为 $\sqrt{\mu}$，假设接入和回传信号在频域上采用矩形脉冲形状（见 [14]）。尽管这很有意思，但优化 μ 超出了本文的范围，将在未来的工作中探讨。

B. 用户有效速率与网络和速率

令 $D_j(t)$ 表示用户设备 j 在时间 t 的业务请求（单位：比特每秒）。从回传的角度看，$T_{i,j}(t) = \min\left(D_j(t)\ C_i^{(a),j}(t)\right)$ 可视为接入链路 $i \to j$ 上的有效数据需求。我们假设同一回传链路 $i \in \mathcal{S}_0$ 上的用户设备共享其容量 $C_i^{(b)}(t)$。因此，令 $\beta_{i,j}(t) \in [0,1]$ 表示分配给与 mIAB节点 i 通信的用户设备 j 的回传容量份额，时间为 t。于是，用户设备 j 从 IAB 站 i 感知到的瞬时有效速率 $R_{i,j}(t)$ 为 $\forall j \in \mathcal{U}(t)$，其表达式如下：

$$R_{i,j}(t) = \begin{cases} \min\left(T_{i,j}(t), \beta_{i,j}(t) z_i(t) C_i^{(b)}(t)\right), \forall i \in \mathcal{S}_0, \\ T_{i,j}(t), \qquad \text{if } i = 0. \end{cases} \tag{8}$$

最终，网络总和速率 $R(t)$ 表示为：

$$R(t) = \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{U}(t)} R_{i,j}(t), \tag{9}$$

III. 问题表述

我们的目标是通过联合解决四个子问题来最大化长期网络和速率：i）确定mIAB节点的最佳位置；ii）确定活动回传链路；iii）将用户与 选定的回传链路相关联；iv）在长期约束和瞬时约束下，为每个已服务的用户设备分配最优的回传容量：

$$\max_{\mathbf{\Psi}(t)} \ \lim_{T \to +\infty} \frac{1}{T} \sum_{t=1}^{T} R(t), \tag{$\mathcal{P}_0$}$$

$$\text{s.t.} \ \lim_{T \to +\infty} \frac{1}{T} \sum_{t=1}^{T} \kappa(t) \geq \kappa_0, \tag{$\mathcal{C}_1$}$$

$$x_{i,j}(t) \in \{0,1\}, \qquad \forall i \in \mathcal{S}, j \in \mathcal{U}(t), \quad (\mathcal{C}_2)$$

$$\sum_{j \in \mathcal{U}(t)} x_{i,j}(t) \leq L_i, \qquad \forall i \in \mathcal{S}, \quad (\mathcal{C}_3)$$

$$\sum_{i \in \mathcal{S}} x_{i,j}(t) \leq 1, \qquad \forall j \in \mathcal{U}(t), \quad (\mathcal{C}_4)$$

$$z_i(t) \in \{0,1\}, \qquad \forall i \in \mathcal{S}_0, \quad (\mathcal{C}_5)$$

$$\sum_{i \in \mathcal{S}_0} z_i(t) \leq M, \tag{$\mathcal{C}_6$}$$

$$\beta_{i,j}(t) \in [0,1], \qquad \forall i \in \mathcal{S}_0, j \in \mathcal{U}(t), \quad (\mathcal{C}_7)$$

$$\sum_{j \in \mathcal{U}(t)} \beta_{i,j}(t) \leq 1, \qquad \forall i \in \mathcal{S}_0, \quad (\mathcal{C}_8)$$

$$\ell_i(t) \in \mathcal{L} \subset \mathbb{R}^2, \qquad \forall i \in \mathcal{S}_0, \quad (\mathcal{C}_9)$$

$$\|\ell_i(t+1) - \ell_i(t)\| \leq \Delta\ell, \qquad \forall i \in \mathcal{S}_0, \ (\mathcal{C}_{10})$$

其中 $\mathbf{\Psi}(t) = \{x_{i,j}(t)\ z_i(t)\ \beta_{i,j}(t)\ \ell_i(t)\ \forall i,\ j\}$ 而（$\mathcal{P}_0$）中的期望是针对随机业务请求和信道实现取的，其统计特性未知。这里，约束（$\mathcal{C}_1$）保证长期波束利用率的最低值；（$\mathcal{C}_2$）-（$\mathcal{C}_4$）限制每个 IAB 站最多同时服务 L_i 个用户设备，且每个用户设备在任一时刻只能关联到一个 IAB 站。此外，（$\mathcal{C}_5$）-（$\mathcal{C}_6$）确保同时激活的回传链路最多为 M 条。另外，（$\mathcal{C}_7$）-（$\mathcal{C}_8$）保证分配给同一回传链路上用户设备的回传容量份额 $\beta_{i,j}(t)$ 为正且每个时刻的总和至多为一 t。最后，（$\mathcal{C}_9$）-（$\mathcal{C}_{10}$）确保 mIAB节点在空间区域 $\mathcal{L}$ 内移动时，每次移动距离不超过 $\Delta\ell$ 米。注意，mIAB节点的位置 $\ell_i(t)$ 会影响路径损耗，进而影响信道增益和 $R(t)$。

问题（$\mathcal{P}_0$）是一种组合非凸优化问题，其复杂度随着用户设备数量的增加而呈指数级增长相对于，因此使用传统优化工具难以求解[15]。此外，上述子问题的解会相互影响。例如，最优回传容量分配取决于用户关联，而用户关联又取决于mIAB节点的位置，这是由于接入与回传链路之间存在相互干扰。

命题 1. 给定*mIAB*节点的部署，若已知最优用户关联，则，

1) 最优回传链路关联为：

$$z_i^*(t) = \mathbb{1}\left(\rho_i(t) > 0\right), \quad \forall i \in \mathcal{S}_0, \tag{10}$$

其中，$\rho_i(t) = \sum_{j \in \mathcal{U}(t)} x_{i,j}^*(t)$ 是*mIAB*节点 i的负载。这里，$\mathbb{1}(\text{cond}(x))$ 是指示函数，当$1\ \text{cond}(x)$成立时取值为0 ，否则取值为

2) 通过求解以下凸优化问题，可获得回传容量的最优分配：

$$\max_{\{\beta_{i,j}\}_{i,j}} \sum_{i \in \mathcal{S}_0^+} \sum_{j \in \mathcal{U}^+} \min\left(T_{i,j}(t), \beta_{i,j}(t) C_i^{(b)}(t)\right), \tag{$\mathcal{P}_1$}$$

$$\text{s.t.} \ \ (\mathcal{C}_7) \ \textit{and} \ (\mathcal{C}_8),$$

其中，$\mathcal{S}_0^+$ 表示具有活动回传链路的*mIAB*节点集合，$\mathcal{U}^+$ 表示活跃用户设备集合。

因此，($\mathcal{P}_0$) 可归结为：*i) mIAB* 节点协调问题，即确定每个时刻*mIAB*的最优位置部署；以及*ii)*用户关联问题，即确定用户设备和*mIAB*的最优分配，以最大化长期网络和速率。

证明概要。 证明首先注意到，若没有用户设备请求连接到相应的mIAB节点，则无需激活回传链路。反之，若回传链路未激活，则无需将用户设备关联到相应的mIAB节点。从接入网络的角度看，这种用户关联是最优的；从回传角度看也同样最优（同时保证约束($\mathcal{C}_6$)）。因此，若已知最优用户关联，则可利用式(10)立即推导出回传链路的最优关联。在这种情况下，给定mIAB节点的部署，($\mathcal{P}_0$)可化简为($\mathcal{P}_1$)：一个仅针对 $\{\beta_{i,j}(t)\ \ \forall i,j\}$的优化问题。随后，观察到约束($\mathcal{C}_7$)和($\mathcal{C}_8$)均为凸约束，即可得出($\mathcal{P}_1$)的凸性。此外，$C_i^{(b)}(t)$和$T_{i,j}(t)$为常数、正值，且与$\beta_{i,j}(t)\ \ \forall i,j$无关，因此$\min\left(T_{i,j}(t)\ \beta_{i,j}(t) C_i^{(b)}(t)\right)$ 是关于$\beta_{i,j}(t)$的凹函数。由此即完成证明。 □

注释 1. 由于我们假设所有毫米波链路之间的频率可完全空间复用，因此变量 $\beta_{i,j}(t)$ 在($\mathcal{P}_1$)中相互解耦，可在每个 *mIAB*节点上分布式求解。

根据命题 1，我们现在的目标是：i) 找出每个时刻 mIAB 的最优位置部署；ii) 确定最优用户

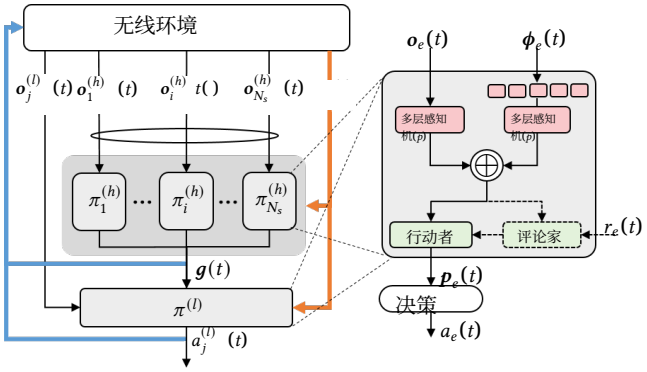

图 2. 分层策略网络架构。

关联，同时考虑接入和回传。由于这两个子问题具有组合性且非凸，联合求解仍然十分复杂。考虑到网络动态（包括可变规模拓扑、阴影衰落、衰落以及用户设备移动），求得最优解更具挑战性。为降低此类复杂性，我们采用以下分层多智能体强化学习方法。

IV. 基于分层多智能体强化学习的解决方案

A. 通用分层优化框架

为求解问题（$\mathcal{P}_0$），我们将标准强化学习框架扩展为分层两级结构，其中高层策略 $\pi^{(h)}$（回传链路管理器）制定 mIAB 节点的位置部署策略，并为低层策略 $\pi^{(l)}$（接入链路管理器）设定目标；低层策略则确定用户关联策略。具体而言，高层策略负责协调 mIAB 节点，动态调整其位置，以共同最大化用户覆盖（至少由一个 mIAB 节点提供覆盖）和回传容量总和 $C^{(b)}(t) = \sum_{i\in\mathcal{S}_0} C_i^{(b)}(t)$。随后，低层策略确定用户关联，以最大化网络和速率（9），同时保证长期波束利用率（$\mathcal{C}_1$）。我们将这两个优化问题分别转化为多智能体强化学习框架，分别将每个 mIAB 和 UE 建模为自主智能体，并通过它们与无线环境的交互分别学习高层和低层策略。在每个时刻 t，无线环境分别向每个 mIAB 节点 i 和每个 UE j提供高层和低层观测 $\mathrm{o}_i^{(h)}(t) \in \mathbb{R}^h$ 和 $\mathrm{o}_j^{(l)}(t) \in \mathbb{R}^l$。在我们的设定中，mIAB 节点 i 的观测与其当前位置一致；因此，$\mathrm{o}_i^{(h)}(t) = \ell_i(t)$。相比之下，我们将 UE j的观测定义为 $\mathrm{o}_j^{(l)}(t) = \{D_j(t)、\mathrm{RSS}_i(t)、\mathrm{AoA}_i(t)\ R_i、_j(t-1)\}_{i\in\mathcal{S}}$，其中包括其业务请求、瞬时局部信号测量值（如接收信号强度（RSS））、相应的估计到达角（AoA），以及相对于不同 IAB 站点先前经历的吞吐量，类似于 [15]。除局部观测外，我们还假设每个实体e（无论是 mIAB 节点还是 UE）都能构建局部无线电地图$\phi_e(t) \in \mathbb{R}^{p\times n}$，以反映其对周围无线环境的相对感知。结合局部观测，该地图有助于有效学习分层策略。

B. 学习局部无线电地图

在我们的框架中，每个实体都学习根据位置信息生成局部无线电地图 $\phi_e(t) \in \mathbb{R}^{p\times n}$ $\{\ell_{e'}(t)\ \ e' \in \mathcal{N}_e(t)\}$ 由邻近实体发出信号 $\mathcal{N}_e(t)$，以限制复杂度。在这个实体数量及其位置随时间变化的动态网络中，$\mathcal{N}_e(t)$的大小和接收消息的顺序也会相应变化。为了构建$\phi_e(t)$，使其相对于 $\mathcal{N}_e(t)$具有大小不变性，并且相对于接收消息具有排列不变性，我们借鉴了神经注意力机制 [16]的思想。具体而言，我们定义 $\mathrm{k}_e^{(d)}, \ell_{e'}(t)^T \in \mathbb{R}^n\ \forall d = 1\ldots N_d$，$\mathrm{w}_e^{(d)}, (\ell_e(t) - \ell_{e'}(t))^T \in \mathbb{R}^n$ 以及 $\mathrm{v}_e^{(d)}, \mathrm{w}_e^{(d)}, (\ell_e(t)(\,)),,$，并将其分别称为实体 e 相对于实体 e'的相对键和值。这里，N_d是滤波器的数量（也称为注意力头 [16]）；$\mathrm{w}_{e\,k}^{(d)},, \mathrm{w}_{e,v}^{(d)} \in \mathbb{R}^{n\times 2}$ 是可学习参数。类似地，令 $\mathrm{q}_e^{(d)} = \mathrm{w}_{e,q}^{(d)}\ell_e(t)^T \in \mathbb{R}^n$ 表示实体 e $\mathrm{q}_e^{(d)} = \mathrm{w}_{e\,q}^{(d)}$ 的查询，$[\mathrm{k}_{e\,e'}^{(d)}\ \ \forall e' \in \mathcal{N}_e(t)]$ 其中 w , 也是可学习参数。令 V v , , ( ) 表示值矩阵。我们使用点积机制 [16]计算注意力矩阵 $\mathrm{A}_e^{(d)}$：$\mathrm{A}_e^{(d)} = \mathrm{softmax}\left(\left[(\sqrt{n})^{-1}\mathrm{q}_e^{(d)}\mathrm{k}_{e,e'}^{(d)T}, \forall e' \in \mathcal{N}_e\right]\right)$，其中 softmax(·) 是归一化指数函数。它表示在第 d个注意力头中，实体 e 与其邻近实体 e' 之间的交互。最后，我们通过拼接各注意力头的输出，计算实体 e 的局部无线电地图；这些输出通过 $\mathrm{A}_e^{(d)}$与 $\mathrm{V}_e^{(d)}$之间的标量积获得：

$$\phi_e(t) = \mathbf{w}_{e,\phi}\left[\bigoplus_{d=1}^{N_d}\mathbf{A}_e^{(d)}\mathbf{V}_e^{(d)}\right]. \tag{11}$$

这里，$\mathbf{w}_{e,\phi} \in \mathbb{R}^{p\times N_d}$ 是一个可学习参数，$\oplus$ 表示行拼接算子。

C. 学习高层策略

在我们的场景中，每个移动式集成接入与回传（mIAB）i 都维护自己的策略 $\pi_i^{(h)}$。给定 $\langle \mathrm{o}_i^{(h)}(t)\ \phi_i(t)\rangle$，$\pi_i^{(h)}$ 会生成概率分布$\boldsymbol{p}_i^{(h)}(t) = \pi_i^{(h)}(\mathrm{o}_i^{(h)}(t)\ \phi_i(t))$，其定义在动作空间 $\mathcal{A}^{(h)}$上，表示沿x轴或 y轴移动的可能方向集合，其中包括静止状态，即 $\mathrm{card}(\mathcal{A}^{(h)}) = 5$。根据$\boldsymbol{p}_i^{(h)}(t)$，mIAB i 会采样高层动作 $a_i^{(h)}(t)$，并沿选定方向以固定步长 $\Delta\ell$ 移动，以最大化时间范围 T_h内的预期 γ_h-折扣回报总和 $\sum_{t=0}^{T_h-1}\gamma_h^t r_i^{(h)}(t)$。因此，我们定义高层奖励，以最大化用户覆盖和总计回传容量：

$$r_i^{(h)}(t) = -(1-\delta_i^{(h)}(t))d_i(t) - \delta_i^{(h)}(t)(d_0 - C^{(b)}(t)). \tag{12}$$

此处，$\delta_i^{(h)}(t) = \mathbb{1}(d_i(t) < d_0)$，其中 d_0 是理想的参考距离，$d_i(t) = \|\ell_i(t) - \ell_i^*(t)\|$ 是当前 mIAB节点 i的位置与其最优位置$\ell_i^*(t)$之间的距离。由于事先无法获知 $\ell_i^*(t)$，我们仅在训练过程中用最近质心的位置对其进行近似，并在线性分配给 mIAB节点 i 之前，先使用例如 K均值算法对用户设备进行聚类。通过这种方式，我们促使 mIAB节点移向能够共同最大化用户覆盖（式 (12) 的第一项）和回传容量（式 (12) 的第二项）的位置。

D. 学习低层策略

与移动集成接入与回传节点不同，用户设备共享同一策略 $\pi^{(l)}$，从而降低了复杂度。给定 $\langle \mathrm{o}_j^{(l)}(t)\ \phi_j(t)\rangle$ 以及目标

$\boldsymbol{g}(t) = \{\ell_i(t) \quad \forall i \in \mathcal{S}_0\}$ 由高层策略定义，$\pi^{(l)}$ 会在动作空间 $\mathcal{A}^{(l)} \subset (\mathcal{S} \cup \{\emptyset\})$ 上生成概率向量 $\boldsymbol{p}_j^{(l)}(t) = \pi^{(l)}(\mathbf{o}_j^{(l)}(t)$ $\phi_j(t)|\boldsymbol{g}(t))$，因为用户设备（例如 处于断连状态时）可能未关联到任何站点。基于 $\boldsymbol{p}_j^{(l)}(t)$，用户设备 j 采样其动作 $a_j^{(l)}(t)$，该动作对应于保持空闲或向某个 IAB 站发送关联请求的决策，以最大化时间范围 T_l内的 γ_l-折扣回报 $\sum_{t=0}^{T_l-1} \gamma_l^t r_j^{(l)}(t)$ 期望总和。我们将目标条件奖励定义为：

$$r_j^{(l)}(t) = (1 - \delta^{(l)}(t))\kappa(t) + \delta^{(l)}(t)(\kappa_0 + R(t)). \qquad (13)$$

与（12）类似，我们定义 $\delta^{(l)}(t) = \mathbb{1}(\kappa(t) \geq \kappa_0)$，以便最大化（13）时能够联合优化波束覆盖率 $\kappa(t)$，从而保证约束（$\mathcal{C}_1$）（（13）的第一项），并最大化网络和速率 $R(t)$（（13）的第二项）。

注 2。 实际上，高层决策仅每隔 T_l 步做出一次，这对应于回传更新频率，并持续 T_h 步；在此之后，低层策略运行以生成低层动作。此外，为了实现有效学习，我们将 (d_i)（见 （12））相对于预定义的最大距离进行归一化，并将 $(C^{(b)})$（见 （12））和 (R)（见 （13））按照其平均值（相对于环境随机性）进行归一化。

E. 策略架构和学习机制

我们对高层和低层策略采用相同的架构，如图 2 所作的简要描述。首先，我们使用具有 p 个神经元的多层感知机（MLP）对每个实体的局部观测和无线电地图（展平后）进行编码。然后，我们将所得的编码向量连接起来，作为演员-评论家框架的输入，并使用广为人知的近端策略优化 [17]对其进行端到端优化。完整描述请参阅 [17]。

V. 数值结果

我们在$N_s = 3$ 个mIAB节点和一个IAB供体的覆盖范围内随机部署 $K_0 = 25$ 个用户设备。在这个动态网络中，用户设备的数量和位置会随时间变化。我们采用随机路点移动模型来描述用户设备的移动（速度为 $\sim [0, 1]$ 米/秒$^{-1}$），这是标准做法 [18]。不过，我们提出的机制在设计上不依赖于用户设备的移动模型。此外，我们将用户设备业务请求的动态变化建模为泊松过程，其强度从三种业务类型中随机选取，对应的平均数据速率需求分别为5 Mbps、200 Mbps和1.5 Gbps。表I总结了仿真参数。我们通过经验确定学习参数 $p = n = 128$，$N_d = 8$，并使用一个包含 $2p$ 个神经元的多层感知机组成行动者和评论家网络。所有层均采用整流线性单元激活函数。此外，我们将行动者和评论家的学习率设为 10^{-4}，折扣因子设为 $\gamma_h = 0.95$、$\gamma_l = 0.6$和 $T_h = T_l = 250$。

分层策略收敛。我们首先评估所提出方案的收敛特性。图3分别展示了训练过程中各回合的高层奖励、低层奖励和波束覆盖率。为清晰起见，我们绘制了在长度为100的窗口上计算的相应滚动平均值和标准差。在图3a中，虚线区域表示$\delta_i^{(h)}(t)$ 等于零的情况。在此区域内，移动集成接入回传代理正在优化

表 I 仿真参数 [3]

参数	值
载波频率	28 GHz
系统带宽，B	300 MHz
带宽划分，μ	1/3
热噪声，N_0	−174 分贝毫瓦/赫兹
小尺度衰落（$\sim m$-Nakagami）	$m = 3$
大尺度衰落（$\sim \mathcal{N}(0, \sigma_0^2)$）	$\sigma_0^2 = 9$ dB（用户）；$\sigma_0^2 = 3$ dB（中继）
路径损耗（中继-用户、宿主-中继）	$132.89 + 25$对数$_{10}(d$ [千米])
路径损耗（宿主-用户）	$154.1 + 25\log_{10}(d$ [公里])
发射功率（回传 / 接入）	43 dBm / 33 dBm
天线增益（IAB节点 + 供体节点）	10x10 天线阵列 [15, 图 5，示意图 2]
天线增益（用户）	0 dBi（全向）
波束成形	$L_i = 10\ \forall i \in \mathcal{S}_0$；$L_0 = 7$；$M = 3$
小区半径	100 m
覆盖范围	50 m（节点）；75 m（供体节点）
步长 $\Delta\ell$	5 m
目标值	$\kappa_0 = 0.8$；$d_0 = 10$ 米
自干扰增益 ξ	−100 dBm

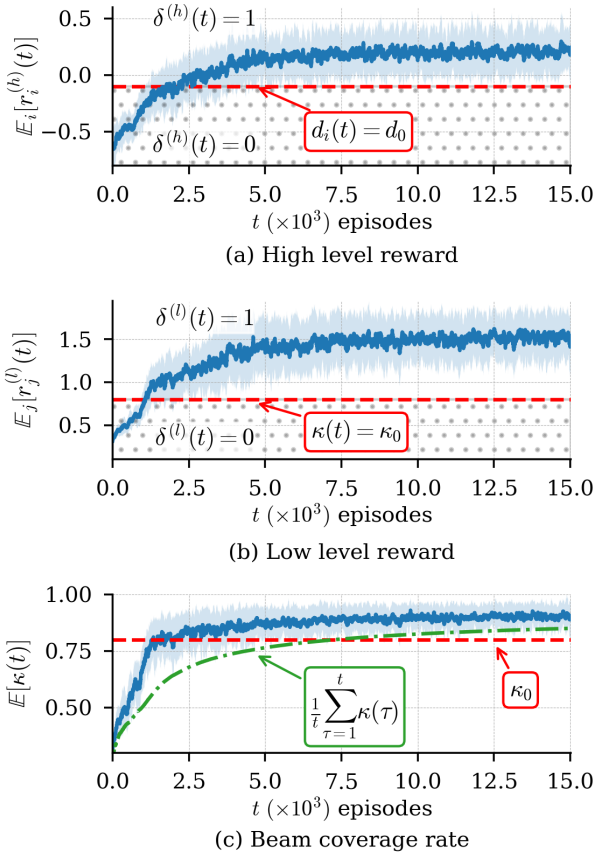

图3. 分层策略学习的收敛性。

用户覆盖（见式（12））。奖励曲线越过该区域，表明高层策略不仅有效地学会了通过调整mIAB节点的位置来优化用户覆盖，还学会了优化回传容量。图3b中的低层策略也是如此，它同样学会了联合优化波束覆盖和网络和速率，因为低层奖励也越过了虚线区域；在该区域中，重点仅在于最大化$\kappa(t)$（见式（13））。图3c进一步证实了这一点，其中我们保证了长期平均约束（$\mathcal{C}_1$）。

与基准方法的性能比较。

图4比较了我们所提出方案的性能

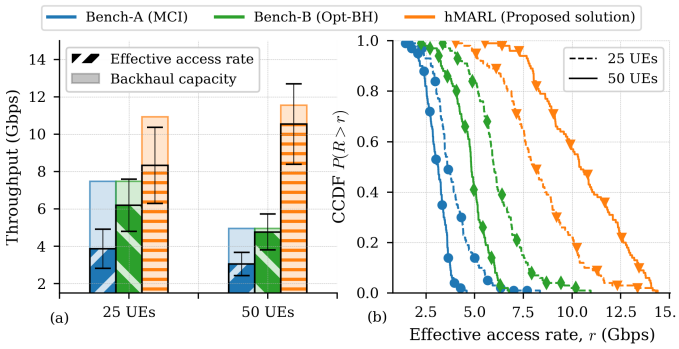

图4。与基准方法的性能比较。我们基于500次随机蒙特卡罗部署和信道实现，绘制直方图及相应的累积分布函数（CCDF）。

（称为 hMARL）与采用最大信噪比算法管理接入链路的两种基准方案：该算法将每个用户设备分配给能够提供最大信噪比（SNR）的 IAB 站。第一种基准方案（称为 Bench-A）采用最大载波干扰（MCI）策略分配回传容量，其中分配给某个用户设备的容量部分与其频谱效率成正比 [5]。第二种基准方案（称为 Bench-B）通过求解我们提出的凸优化问题（$\mathcal{P}_1$）来最优分配回传容量。所有这些基准方案都采用集中式穷举 K均值聚类算法，将每个 mIAB节点部署在最近聚类的质心处。相比之下，我们提出的方案能够自主且分布式地部署 mIAB节点，以最大化用户覆盖和回传容量，同时优化用户关联，而无需任何中央协调器。为了展示所提方案的有效性，我们重点关注 mIAB 网络，因为与 IAB供体关联的用户设备不依赖回传链路。因此，从图4a的直方图中可以观察到，我们提出的方案明显优于两种基准方案。与 Bench-A 和 Bench-B 相比，该方案使回传容量额外提升了 50%，并分别将网络和速率提高了 115% 和 34.4%。这一显著增益在图4b中同样十分明显，该图绘制了服务覆盖概率，即，即网络和速率超过给定阈值的概率。当我们将该阈值设置为例如6.5 Gbps（对应 $K = 25$ 个用户设备）时，我们提出的方案可保证 84% 的服务覆盖率，而 Bench-A 和 Bench-B 分别仅为 3% 和 40%。此外，尽管我们针对 $K = 25$ 个用户设备进行了训练，但我们还评估了在 $K = 50$ 情况下、无需重新学习过程时的性能，以展示所提方法应对不同用户设备数量和网络拓扑的能力。随着用户设备数量增加，基准方案的性能相应下降。事实上，由于同信道干扰和有限的回传容量，网络和速率不会随着用户设备数量的增加而系统性提升。然而，与基准方案不同，我们提出的方案仍表现出额外的性能提升，其网络和速率最多达到 Bench-A 和 Bench-B 的 $3.5\times$ 和 $2.2\times$ 。这一显著增益源于我们算法能够根据用户设备的业务请求为其提供适当服务，并相应地调整回传网络。

VI. 结论

本文研究了移动集成接入与回传节点定位、用户关联和回传容量分配的优化问题。我们聚焦于带内中继IAB网络，该网络在同一频段上承载接入与回传链路。这类网络会受到严重的干扰约束，主要限制回传容量和终端用户的性能。为解决这一问题，我们提出了一种新颖且可扩展的两级分层多智能体学习机制，其中高层策略通过联合优化用户覆盖和回传容量来确定移动集成接入与回传节点的位置。该过程为低层策略设定目标，由低层策略优化用户关联并分配回传容量。完成学习后，这些策略将被分发，并能够根据相对于环境动态和接入网络需求自主重新配置回传网络。数值评估表明，与基准方法相比，我们的方案可使网络和速率提升至多 $3.5\times$ 。未来的工作将利用本研究的成果，为无人机网络提供回传支持。

参考文献

[1] 3GPP TR 38.913，“第五代移动通信技术；下一代接入技术的场景与需求研究（第15版）”，2018年9月。[2] D. Lopez-Perez、I. Guvenc 等，“异构网络中增强小区间干扰协调面临的挑战”，IEEE无线通信，第18卷，第3期，第22–30页，2011年。[3] 3GPP TR 38.874，“集成接入与回传研究（第16版）”，2018年12月。[4] R. J. Weiler、M. Peter 等，“利用毫米波实现第五代移动通信技术回传与接入”，载于《欧洲网络与通信会议》（EuCNC）论文集，第1–5页，2014年6月。[5] A. De Domenico 等，“异构蜂窝网络的回传感知小区选择算法”，载于IEEE个人、室内和移动无线电通信国际研讨会（PIMRC）论文集，第1688–1693页，2013年9月。[6] M. Polese、M. Giordani 等，“毫米波集成接入与回传的分布式路径选择策略”，载于IEEE全球通信大会（GLOBECOM）论文集，第1–7页，2018年。[7] E. Pateromichelakis 和 K. Samdanis，“面向毫米波回传/接入网络的上下文感知联合路由与调度”，载于IEEE全球通信大会（GLOBECOM）论文集，第1–6页，2018年。[8] D. Yuan、H.-Y. Lin 等，“毫米波蜂窝网络中的最优联合路由与调度”，载于IEEE计算机通信会议（INFOCOM）论文集，第1205–1213页，2018年。[9] P. Valente Klaine、M. Jaber 等，“基于Q学习的回传感知用户专属小区关联”，IEEE无线通信汇刊，第18卷，第7期，第3528–3541页，2019年。[10] W. Lei、Y. Ye 等，“集成接入与回传网络中的基于深度强化学习的频谱分配”，IEEE认知通信与网络汇刊，第6卷，第3期，第970–979页，2020年。[11] B. Zhang 和 I. Filippini，“基于多智能体强化学习的毫米波集成接入与回传网络移动性感知资源分配”，载于IEEE移动自组网与智能系统国际会议（MASS）论文集，第17–26页，2021年。[12] A. Fouda 等，“面向第五代移动通信的基于无人机的带内集成接入与回传”，载于IEEE车辆技术会议秋季会议论文集，第1–5页，2018年。[13] R. S. Sutton 和 A. G. Barto，《强化学习：导论》，麻省理工学院出版社，第2版，2018年。[14] A. AlAmmouri 等，“蜂窝网络中的带内 α双工方案：随机几何方法”，IEEE无线通信汇刊，第15卷，第10期，第6797–6812页，2016年。[15] M. Sana、A. De Domenico 等，“动态毫米波网络中用于自适应用户关联的多智能体强化学习”，IEEE无线通信汇刊，第19卷，第10期，第6520–6534页，2020年。[16] A. Vaswani 等，“Attention is All You Need”，载于《神经信息处理系统进展》（NeurIPS）论文集，第5998–6008页，2017年。[17] J. Schulman 等，“近端策略优化算法”，载于CoRR，第abs/1707.06347卷，2017年。[18] D. Mitsche、G. Resta 和 P. Santi，“具有节点均匀空间分布的随机航点移动模型”，《无线网络》，第20卷，第5期，第1053–1066页，2014年。