

基于非稳态 MAB 的 LEO 卫星跳波束时隙分配算法

林敏¹, 阚鹏程¹, 赵柏¹, 程铭¹, 杨绿溪²

(1. 南京邮电大学通信与信息工程学院, 江苏 南京 210003; 2. 东南大学信息科学与工程学院, 江苏 南京 211189)

摘要: 针对低地球轨道 (LEO) 卫星系统中的跳波束资源分配算法不能适应小区业务动态变化等问题, 提出了一种基于非稳态多臂赌博机 (MAB) 的 LEO 卫星跳波束时隙分配算法。首先, 以系统二阶差分容量最小化为优化目标, 建立了时隙分配和波束等级匹配的联合优化问题。其次, 由于该问题非凸且难以直接求解, 基于有效小区和有效关键小区的概念提出波束等级组合方案生成算法, 从而生成所有可能的波束等级组合方案。接下来, 提出了基于非稳态 MAB 模型的动态时隙分配方案, 在最优波束等级组合方案下完成时隙分配与波束等级匹配的联合优化。最后, 计算机仿真结果表明, 所提算法在多种小区业务分布的情况下, 系统平均冗余度均不超过 20%; 相比于其他对比方案, 所提算法在保持较高的系统吞吐量的同时, 还可以将波束平均重访时间控制在 300 ms 左右。

关键词: 低地球轨道; 跳波束; 分等级波束; 非稳态多臂赌博机

中图分类号: TN92

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2023160

Time-slot allocation algorithm for LEO satellite beam hopping based on non-stationary MAB

LIN Min¹, KAN Pengcheng¹, ZHAO Bai¹, CHENG Ming¹, YANG Lyuxi²

1. School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

2. School of Information Science and Engineering, Southeast University, Nanjing 211189, China

Abstract: Aiming at the problem of inadequate adaptation to dynamic changes in cell services in the resource allocation algorithm for hopping beams in low earth orbit (LEO) satellite systems, a time-slot allocation algorithm for LEO satellite beam hopping based on a non-stationary multi-armed bandit (MAB) was proposed. Firstly, the joint optimization problem of slot allocation and beam grading matching was established, with the minimization of the system's second-order differential capacity as the optimization objective. Secondly, due to the non-convexity and difficulty in direct solution of this problem, a beam grading combination scheme generation algorithm was proposed based on the concept of effective cells and effective critical cells, which generated all possible beam grading combination schemes. Next, a dynamic slot allocation scheme based on the non-stationary MAB model was proposed, and joint optimization of slot allocation and beam grading matching was completed under the optimal beam grading combination scheme. Finally, the computer simulation results show that the average redundancy of the proposed algorithm is less than 20% in the case of multiple cell service distributions. In addition, compared with other schemes, the proposed algorithm can control the average beam revisit time to about 300 ms while maintaining high system throughput.

Keywords: LEO, beam hopping, graded beam, non-stationary MAB

收稿日期: 2023-05-09; 修回日期: 2023-08-16

基金项目: 国家自然科学基金资助项目 (No.61971128, No.62301282); 南京邮电大学引进人才科研启动基金资助项目 (No.NY220111, No.NY221009); 江苏省研究生科研与实践创新计划基金资助项目 (No.KYCX22_0965)

Foundation Items: The National Natural Science Foundation of China (No.61971128, No.62301282), Talent Research Start-up Foundation of Nanjing University of Posts and Telecommunications (No.NY220111, No.NY221009), The Postgraduate Research and Practice Innovation Program of Jiangsu Province (No.KYCX22_0965)

0 引言

针对 6G 网络中“泛在连接”的技术愿景，卫星通信因其通信覆盖范围大、不受地理条件限制、组网灵活等优点，被产业界和学术界认为是实现 6G 技术愿景的一种重要的通信方式^[1]。在卫星通信中，低地球轨道（LEO, low earth orbit）卫星通信凭借其传输时延短、建设成本相对较低等特点，在近几年得到了较多的关注和发展^[2]。随着 SpaceX、OneWeb 等公司低轨卫星星座建设计划的提出与实施，以及我国低轨卫星互联网星座的研究和部署，LEO 卫星通信成为卫星通信领域研究的焦点。

跳波束（BH, beaming hopping）技术以业务需求为驱动，通过在同一时刻只激活部分波束覆盖相应的地面小区，实现系统资源的合理分配和灵活调度，一经提出便得到了广泛的关注和研究^[3-4]。文献[5-6]综合考虑了跳波束系统中波束重访时间和分簇尺寸等因素，以 n 阶差分容量最小化为目标建立相应的优化问题，采用凸优化方法求解得到时隙分配结果，并通过时隙分配进一步缓解了多波束卫星的同频干扰问题。然而，该方法采用离线静态资源分配，并不能很好地适应用户业务需求动态变化的场景。文献[7]则基于贪婪算法根据用户业务的需求分布灵活地分配星载资源，但是该算法未考虑服务公平性问题。文献[8]则在保证每个波束服务公平性的前提下，建立了最小化实时性服务数据包时延以及最大化非实时性服务数据包吞吐量的多目标优化问题，并基于深度强化学习实现了该多目标优化问题的动态求解。该方法虽然能很好地适应用户业务需求动态变化的场景，但仍然存在学习模型复杂度较高、对计算资源需求较高的问题。

多臂赌博机（MAB, multi-armed bandit）模型作为强化学习中实现探索和利用平衡的一类经典模型，因其具有在未知环境信息条件下也能实现动态优化的特点，在信道选择、推荐系统等领域得到了广泛的应用^[9-11]。传统的 MAB 模型依赖于平稳的奖励分布，文献[12]进一步研究了非稳态 MAB 模型并提出了相应的求解算法，并且通过理论和实践证明，该求解算法在保证收敛性的情况下很好地应用于奖励分布未知且随时间动态变化的场景，同时相比于深度强化学习等人工智能算法，其对计算资源的需求更低。

本文考虑 LEO 跳波束卫星系统中地面小区业务需求分布不均以及动态变化的场景，在满足小区业务需求量的基础上建立以二阶差分容量为目标

的最小化问题。由于小区业务需求的时变性，所建立的优化问题难以直接求解，因此本文首先提出波束等级组合方案生成算法以生成所有可能的波束等级组合方案；然后在固定波束等级组合方案的情况下，通过非稳态 MAB 模型完成时隙的动态分配；最后以二阶差分容量最小化为准则选择最优波束等级组合方案，从而完成时隙分配与波束等级匹配的联合优化。由仿真结果可知，与单一等级方案和单一波束方案相比，本文算法能够进一步降低跳波束系统业务处理的归一化平均冗余度；同时相比于其余的时隙分配方案，本文算法在总业务需求量变化的同时能达到较高的系统吞吐量，并保持相对较低的波束平均重访时间。

1 系统模型

如图 1 所示，本文研究 LEO 跳波束卫星通信系统的前向链路传输场景，它由网络控制中心、配备跳波束控制器的 LEO 卫星和地面小区组成，并采用兼容 DVB-S2/S2X 协议的跳波束工作方式。综合考虑 LEO 卫星所处的轨道高度以及移动性，本文将 LEO 卫星在飞行过程中的总体覆盖区域划分为不同的观察窗口，并且在每个观察窗口内能够服务 N 个地面小区。为了提高系统的频谱利用效率，LEO 卫星在当前观察窗口内以全频率复用的方式同时生成 K 个点波束服务 N 个小区。由于星载资源有限，在跳波束卫星通信系统中通常假设点波束数目远小于小区数目，即 $K \ll N$ 。基于上述的 LEO 卫星跳波束通信场景，本文将分别详细介绍小区业务模型、信道模型以及分等级波束模型。

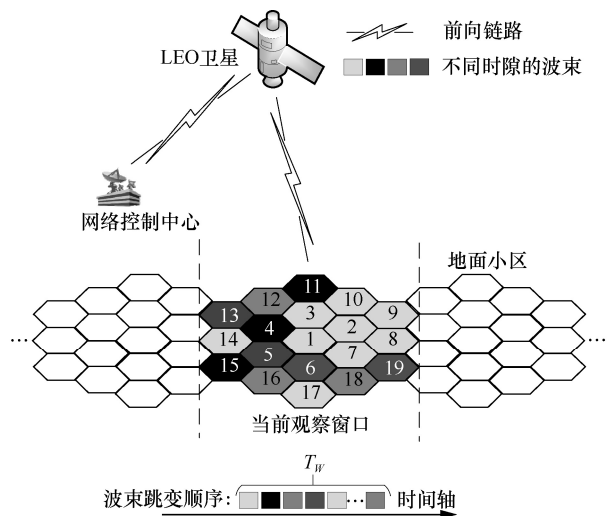


图 1 LEO 跳波束卫星通信系统模型

假设跳波束卫星通信系统的跳波束周期为 T_w ，而卫星在跳波束周期内的每个时隙以数据包的形式向地面小区提供业务数据，且数据包大小均为 X bit/s。同时假设各个小区在 t 时隙到达的数据包个数为 $A(t) = [a_1(t), a_2(t), \dots, a_N(t)]^T$ ，其中数据包的到达数量 $a_n(t)$ 服从均值为 $\lambda_n (n=1, 2, \dots, N)$ 的泊松分布，则在该观察窗口中各个小区在一个跳波束周期内的业务需求量为

$$R_n = X \sum_{t=1}^{T_w} a_n(t), n=1, 2, \dots, N \quad (1)$$

此外，各小区数据包达到后若不能及时处理则进入缓冲区队列等待，并将 t 时隙的缓冲区数据量表示为

$$\Phi(t) = [\varphi_1(t), \varphi_2(t), \dots, \varphi_N(t)]^T \quad (2)$$

在传统的跳波束卫星通信系统中，其单波束业务数据处理能力较固定，这可能会导致在跳波束时隙分配过程中存在以下 2 种固有问题。1) 当某些小区的业务需求量较大以至于接近甚至超过了单波束最大业务数据处理能力时，便需要卫星波束长期照射这些小区，这会导致其他小区业务处理不及时、数据包等待时延过长；2) 当某几个相邻小区的总业务需求量较低时，如果仍然使用单一等级波束在这些小区间进行跳变，则会占用过多的跳波束时隙，导致其余小区的业务需求量难以得到满足。针对上述问题，本文采用波束分级方法来进行提高跳波束 LEO 卫星通信系统的资源利用率^[13-14]。具体来讲，本文考虑采用 3 种等级的波束处理小区的业务数据包，如图 2 所示。

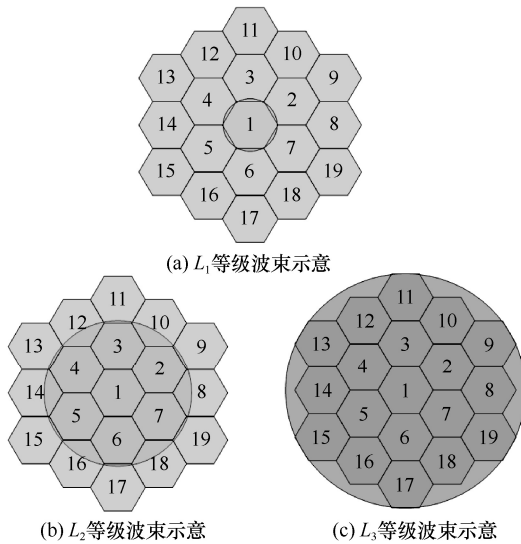


图 2 分等级波束示意

本文考虑 3 种等级的波束，并分别记为 L_1 、 L_2 和 L_3 。其中， L_1 等级波束覆盖 1 个小区， L_2 等级波束覆盖 7 个小区， L_3 等级波束覆盖 19 个小区。综合考虑分等级波束增益差异性、卫星信号传播过程中的自由空间传输损耗以及雨衰等因素的影响，卫星下行信道可建模为^[14]

$$h_{kn, L_i} = \sqrt{G_{kn, L_i}^{tx} G_n^{rx}} \left(\frac{c}{4\pi d_n f} \right) r_n^{\frac{1}{2}} \quad (3)$$

其中， h_{kn, L_i} 表示采用 L_i 等级的第 k 号波束到小区 n 的信道参数； $\frac{c}{4\pi d_n f}$ 表示自由空间传播损耗，且 c 、 f 和 d_n 分别表示光速、载波频率和小区 n 到卫星的距离； r_n 表示雨衰系数，其 dB 形式服从对数正态随机分布，即 $\ln(r_n^{dB}) \sim \text{CN}(\mu_s, \sigma_s^2)$ ，其中 μ_s 和 σ_s 取决于卫星的通信频率、极化方式和用户的位置； G_{kn, L_i}^{tx} 表示采用 L_i 等级的第 k 号波束到小区 n 的发射天线增益，其具体表达式为

$$G_{kn, L_i}^{tx} = G_{\max, L_i} \left(\frac{J_1(\delta_{kn, L_i})}{2\delta_{kn, L_i}} + 36 \frac{J_3(\delta_{kn, L_i})}{\delta_{kn, L_i}^3} \right)^2 \quad (4)$$

其中， $\delta_{kn, L_i} = 2.07123 \frac{\sin \theta_{kn}}{\sin \theta_{3dB, L_i}}$ ， θ_{kn} 为波束 k 的波束中心与小区 n 关于卫星连线之间的夹角， $J_1(\cdot)$ 和 $J_3(\cdot)$ 分别为第一类一阶和第一类三阶贝塞尔函数。根据天线增益相关理论，随着波束覆盖范围的增加，其峰值发射增益下降。具体来讲，若以半功率波束宽度表示波束的有效覆盖范围，则峰值发射增益与半功率波束宽度的关系可表示为^[15]

$$G_{\max, L_i} = \frac{28000}{(\theta_{3dB, L_i})^2} \quad (5)$$

其中， θ_{3dB, L_i} 为 $L_i (i=1, 2, 3)$ 等级波束对应的半功率波束角度，其具体值可通过空间几何关系得到。例如，用 L_i 等级波束覆盖 1 号小区，并假设卫星高度为 H ，小区边长为 d ，如图 3 所示。

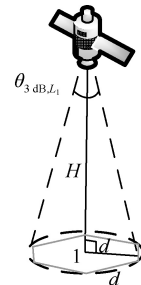


图 3 波束 3 dB 角度计算示意

通过几何推导，可得出 $\theta_{3\text{dB},L_1}$ 具体表达式为

$$\theta_{3\text{dB},L_1} = 2 \arctan \frac{d}{H} \quad (6)$$

同理，可以得出 $\theta_{3\text{dB},L_2}$ 和 $\theta_{3\text{dB},L_3}$ 具体表达式分别为

$$\theta_{3\text{dB},L_2} = 2 \arctan \frac{\sqrt{7}d}{H} \quad (7)$$

$$\theta_{3\text{dB},L_3} = 2 \arctan \frac{\sqrt{19}d}{H} \quad (8)$$

根据式(3)可以得出 t 时隙被 L_i 等级波束 k 服务的小区 n 的输出信干噪比 (SINR, signal to interference plus noise ratio) 表达式为

$$\gamma_{kn,L_i}(t) = \frac{P_T |h_{kn,L_i}|^2}{I_n(t) + \sigma^2} \quad (9)$$

其中， P_T 表示卫星发射功率； $\sigma^2 = \kappa T^{\text{noise}} W$ 表示噪声功率，其中 κ 、 T^{noise} 和 W 分别表示玻耳兹曼常数、噪声温度和载波带宽； $I_n(t)$ 表示该小区在当前时隙受到来自其他被照射小区的同频干扰，其具体表达式为

$$I_n(t) = \sum_{k' \in K \setminus \{k\}} \sum_{n' \in N_{k'}} |h_{k'n,L_i}(t)|^2 P_T \quad (10)$$

其中， K 为当前工作的波束集合， $N_{k'}$ 为波束 k' 服务的小区集合。进一步地，可以得到 t 时隙针对被照射小区 n 的数据包处理能力为

$$C_{n,L_i}(t) = W \ln(1 + \gamma_{kn,L_i}(t)) \quad (11)$$

于是小区 n 经过一个跳波束周期后的总业务数据处理量为

$$C_n = I_n^{T_W \times 1} \sum_{t=1}^{T_W} \left[\frac{C_{n,L_i}(t)}{X} \right], n=1, 2, \dots, N \quad (12)$$

其中， $\lfloor \cdot \rfloor$ 表示向下取整； $I_n^{T_W \times 1}$ 表示小区 n 的跳波束时隙分配矩阵，当小区 n 在 t 时隙被照射时， $\lfloor I_n^{T_W \times 1} \rfloor_t = 1$ ，否则 $\lfloor I_n^{T_W \times 1} \rfloor_t = 0$ 。

2 问题建立与求解

2.1 问题建立

由小区业务模型和分等级波束模型可知，当部分小区业务数据量需求量较大时，应尽量采用 L_1 等级波束以尽可能满足其业务需求；当部分小区业务数据量较小时，可考虑优先采用 L_2 等级波束或 L_3 等

级波束对这部分小区进行合并，从而将更多的时隙资源分配给其余业务需求量较大的小区。因此，需要通过联合优化时隙分配与波束等级匹配以尽可能满足各小区在一个跳波束周期内的业务需求，本文以二阶差分容量最小化为目标函数，建立如下形式的联合优化问题

$$\begin{aligned} \min_{I_n, L_i^n} & \sum_{n=1}^N (R_n - C_n)^2 \\ \text{s.t.} & \text{C1: } L_i^n \in \{L_1, L_2, L_3\} \\ & \text{C2: } I_n(t) \in \{0, 1\}, t=1, 2, \dots, T_W \\ & \text{C3: } \sum_{n=1}^N \frac{1}{l_n^i} I_n(t) \leq K, t=1, 2, \dots, T_W \end{aligned} \quad (13)$$

其中，C1 限定了每个小区的波束等级选取范围；C2 限定了时隙分配矩阵中每个元素的取值范围；C3 中 l_n^i 表示小区 n 选择的 L_i 等级波束所覆盖的小区数目，因此 C3 表示在同一时隙服务地面小区的波束数不能超过多波束卫星的最大波束数目。

由于该问题为非线性整数规划问题，且由于小区业务需求的动态变化，直接求解该问题将变得十分困难。因此本文首先提出波束等级组合方案生成算法以生成所有可能的波束等级组合方案；然后在固定波束等级组合方案的情况下，通过非稳态 MAB 模型完成时隙的动态分配；最后根据二阶差分容量最小化为准则选择最优波束等级组合方案，从而完成时隙分配和波束等级匹配的联合优化。

2.2 波束等级组合方案生成算法

由于跳波束周期通常较短，且 LEO 卫星过境时间有限，因此为了避免 LEO 卫星波束等级的频繁切换，本文假设在一个跳波束周期中，每个小区只可以固定选择一种等级的波束。基于该假设，本文提出一种波束等级组合方案的生成算法以生成各个小区在一个跳波束周期内所有可能的波束等级组合，从而为后续的动态时隙分配提供先验知识。

首先考虑 2 种特殊波束等级组合方案，即单一等级方案和单一波束方案。在这 2 种方案中，各个小区在一个跳波束周期内均采用同一种波束等级 (L_1 等级或 L_3 等级)，不存在多种波束等级并存的场景，此时，退化为传统的跳波束时隙分配问题或单波束场景。

然后考虑更一般的情况，即部分小区选择 L_2 等级波束，而部分小区选择 L_1 等级波束。此时对波束等级组合方案的分析将变得复杂，因为尽管在一个

跳波束周期中各个小区只能选择一种等级的波束，但是小区对 L_2 等级波束的选择可能在不同时隙上存在重叠，如图4所示。

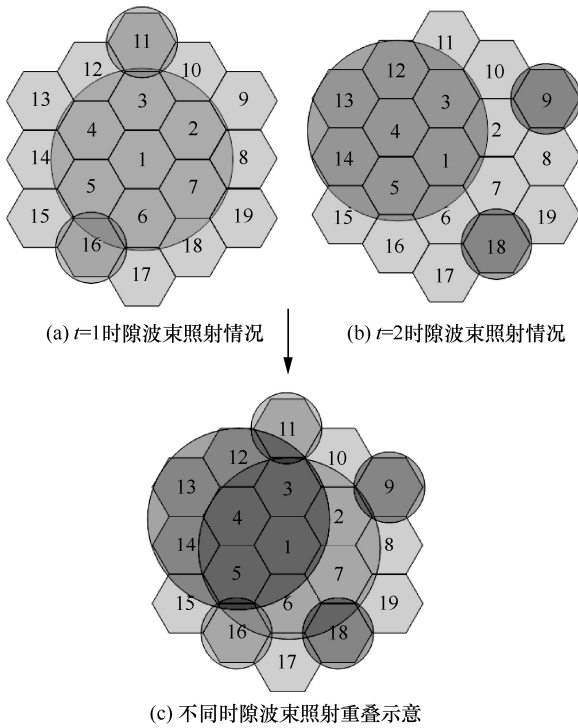


图4 混合波束等级示意

从图4可知，从单个时隙来看，当1~7号小区选择 L_2 等级波束时，其余小区均只可以选择 L_1 等级波束。从整个跳波束周期来看，12~14号小区对应的波束等级依然有可能为 L_2 等级波束，因为它们对应的波束与1~7号小区对应的波束并不在同一个时隙被照射。因此，该波束等级组合方案依然是可行的。波束等级组合方案生成算法首先要处理混合波束等级在不同时隙下的重叠问题。对此，本文首先提出关键小区和有效关键小区的概念。

关键小区为可能处于波束覆盖中心且波束等级的选择将影响周围小区波束等级选择的小区。以图4为例，当1号小区在一个跳波束周期内对应的波束等级为 L_2 等级时，可以通过 L_2 等级波束覆盖方式得出，其周围2~7号小区将至少有3个小区选择 L_2 等级波束，并且1号小区可以处在波束覆盖中心，因此认为1号小区为关键小区。同理，2~7号小区均可作为关键小区。

有效关键小区为在一个跳波束周期中能够处于波束覆盖中心的关键小区。仍然以图4为例，尽管1~7号小区为关键小区，但是在图4所示的波

束等级组合方案中，有效关键小区只有1号小区和4号小区，因为它们在一个跳波束周期中能够处于波束覆盖中心。

通过有效关键小区的概念可知，图4中重叠问题来自在一个跳波束周期中选择了2个关键小区，同时也可以借此计算出混合了 L_2 等级波束的所有可能的波束等级组合方案的数量。具体来讲，当有效关键小区数量为1时，可能的波束组合方案数量为 $C_7^1 = 7$ ，对应的7种方案分别是1~7号小区各自作为选择 L_2 等级波束的有效关键小区，而其余小区选择 L_1 等级波束。因此通过改变有效关键小区的数量，可以得出混合了 L_2 等级波束的所有可能的波束组合方案的数量应为 $\sum_{i=1}^7 C_7^i$ 。

本文提出的波束等级组合方案生成算法的基本流程如下。首先，确定有效关键小区集合 C 以及集合元素个数 M ；然后，依次从集合 C 中选取1个元素、2个元素等，直到 M 个元素全部被取出，并根据每次选取结果确定其余小区的波束等级组合；最后，将所有波束等级组合汇总，得出所有的波束组合方案集合 S 。需要说明的是，尽管前文在引入关键小区时以混合了 L_2 等级波束小区的情况为例，但是通过改变有效关键小区集合 C 中的元素，该算法仍然适用于前文所述的2种特殊波束组合方案的生成。该算法流程如算法1所示。

算法1 波束等级组合方案生成算法

初始化 有效关键小区集合 C ，集合 C 元素数量 M ，波束等级组合方案集合 S

for $i = 1, 2, \dots, M$:

for $j = 1, 2, \dots, C_M^i$:

获取当前已选定有效关键小区组合 $\text{Comb}_{i,j}$;

for $c \in \text{Comb}_{i,j}$:

1) 设定小区 c 波束等级;

2) 若小区 c 波束等级为 L_2 或 L_3 ，则进一步设定其相邻受影响小区波束等级为小区 c 波束等级;

3) 若存在剩余小区波束等级未赋值，则均设定为 L_1 等级;

4) 将当前各小区波束等级组合方案 s 保存到集合 S 中。

end for

end for

end for

2.3 动态时隙分配算法

当生成波束等级组合方案集合 S 后, 本节将针对集合 S 中的每个具体波束等级组合方案 s 进行跳波束时隙分配。由于小区业务需求量是动态变化的, 因此传统的凸优化求解方法将不再适用。本文提出基于非稳态 MAB 模型的求解算法。与传统 MAB 模型类似, 在非稳态 MAB 模型的动作空间中也包含多个摇臂, 而玩家每次只能从动作空间选择一个摇臂, 并获得相应的奖励。因此, 在正式求解非稳态 MAB 模型之前, 应首先定义该模型的动作空间以及奖励设置, 具体如下。

1) 动作空间

当固定一种波束等级组合方案后, 动作空间为该方案下所有跳波束图案构成的集合, 每种跳波束图案对应该动作空间中的一个动作 (摇臂)。以图 5 为例, 假设本文选择的波束等级组合方案为 1~7 号小区在整个跳波束周期内选择 L_2 等级波束, 而其余小区选择 L_1 等级波束, 同时假设每个时隙可同时服务的最大波束数量 $K=3$, 则图 5(a)~图 5(c)均为该波束等级组合方案下可能的 3 种跳波束图案, 该波束等级组合方案下所有可能的跳波束图案数量应为 C_{12}^3 种, 即从 8~19 号小区中任选 2 个 L_1 等级波束小区与 1~7 号小区相组合的组合数量。

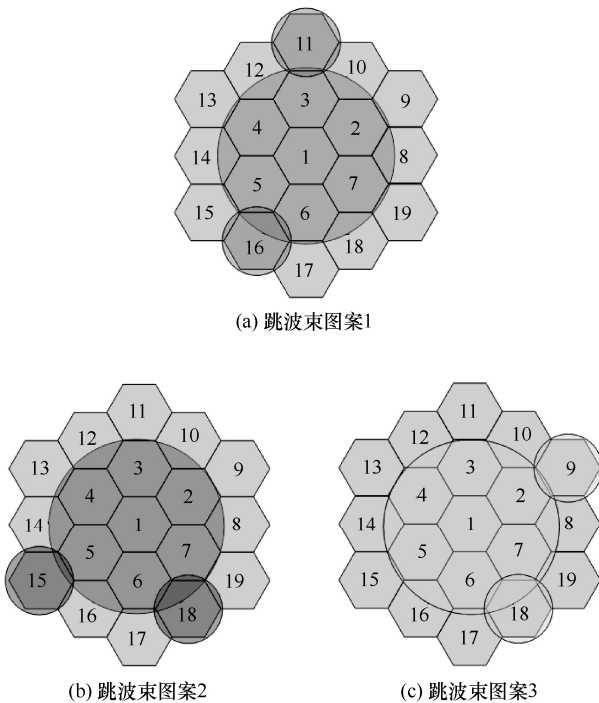


图 5 动作空间示例

2) 奖励设置

奖励设置直接影响非稳态 MAB 模型的决策效果。考虑到跳波束时隙分配的基本要求是尽可能满足每个小区的实际业务需求, 因此本文的奖励设置将围绕小区实际业务需求量和数据包处理量来展开。具体来讲, 假设所选定的动作空间中包含 Q 个摇臂, 且在 t 时隙选择了第 q 个摇臂, 且该摇臂中包含的小区 n 的当前缓冲区数据包个数为 $\varphi_n(t)$, 同时该小区新到达数据包个数为 $a_n(t)$, 且在当前摇臂下计算出该小区的数据包处理能力为 C_n^q , 则针对该小区的瞬时奖励的具体表达式为

$$\text{rwd}_n(t) = \begin{cases} \beta C_n, & (\varphi_n(t) + a_n(t)) \geq 2C_n \\ C_n, & iC_n \leq (\varphi_n(t) + a_n(t)) \leq 2C_n \\ (\varphi_n(t) + a_n(t)) - C_n, & (\varphi_n(t) + a_n(t)) < C_n \end{cases} \quad (14)$$

其中, β 为大于 1 的乘性系数; $C_n = \min(C_n^1, C_n^2, \dots, C_n^Q)$ 为该小区在当前动作空间下的参考数据包处理能力。同时, 由于单个摇臂通常包含多个小区, 因此需要进一步对这些小区的瞬时奖励求平均以得到该摇臂的归一化动作奖励, 则第 q 个摇臂在 t 时隙的奖励表达式为

$$\text{Rwd}_q(t) = \frac{\sum_{n=1}^{\text{Num}_q} \text{rwd}_n(t)}{\text{Num}_q}, q=1, 2, \dots, Q \quad (15)$$

其中, Num_q 表示该摇臂包含的小区数量。

3) 非稳态 MAB 模型的求解

首先, 给定动作空间 A , 该动作空间包含的摇臂数量为 Q_A , 则基于前文的奖励设置, 通过折扣上界置信区间 (D-UCB, discounted upper confidence bound) 算法进行非稳态 MAB 决策, 从而得出该动作空间下的时隙分配方案^[12]。

具体来讲, 首先, 在 t 时隙计算出该动作空间下第 q 个摇臂的折扣期望奖励为

$$\overline{\text{Rwd}}_q(\gamma, t) = \frac{1}{N_q(\gamma, t)} \sum_{s=1}^t \gamma^{t-s} \text{Rwd}_q(s) \mathbf{1}_{\{I_s=q\}} \quad (16)$$

其中, $N_q(\gamma, t) = \sum_{s=1}^t \gamma^{t-s} \mathbf{1}_{\{I_s=q\}}$ 为摇臂 q 截止到 t 时隙的折扣执行总次数; $\gamma \in (0, 1)$ 为预先设定的折扣因子, 表示对历史奖励的利用程度; $\mathbf{1}_{\{I_s=q\}}$ 为指示函数,

若 s 时隙执行动作 q ，该函数值为 1，否则为 0。然后，按照式(17)计算折扣探索因子

$$e_q(\gamma, t) = 2\zeta \sqrt{\frac{\xi \log n(\gamma, t)}{N_q(\gamma, t)}} \quad (17)$$

其中， ζ 与 ξ 均为预先设定的乘性系数， $n(\gamma, t) = \sum_{q \in Q} N_q(\gamma, t)$ 为当前所选动作空间中所有动作的累计折扣执行总次数。之后，将折扣期望奖励与折扣探索因子相加即可得出当前时隙各动作的 D-UCB 索引值 $G_q(\gamma, t)$ 。最后，根据最大索引值选择下一个时隙所要执行的动作。D-UCB 算法如算法 2 所示。

算法 2 D-UCB 算法

```

for  $t = 1, 2, \dots, T_w$  :
    if  $t \leq Q$  :
        执行动作  $I_t = q$  ;
    else
        执行动作  $I_t = \arg \max_{1 \leq i \leq Q} G_q(\gamma, t-1)$  ;
    end if
    计算  $\overline{\text{Rwd}}_q(\gamma, t)$  与  $e_q(\gamma, t)$  ;
    更新索引值  $G_q(\gamma, t) = \overline{\text{Rwd}}_q(\gamma, t) + e_q(\gamma, t)$  ;
end for
    
```

2.4 最优波束等级组合方案选择

利用 D-UCB 算法可得出给定动作空间下的最优时隙分配方案。在这之前，需要选择最优动作空间以确定最优波束等级组合方案。本节考虑以二阶差分容量最小化为准则选择最优波束等级组合方案。具体流程如下。首先，从波束等级组合方案集合 S 中选择一种波束等级组合方案 s ；然后，基于非稳态 MAB 模型完成该方案下的时隙分配；接着，根据时隙分配结果计算方案 s 下小区业务需求量和实际处理量的二阶差分容量，即 $y_s = \sum_{n=1}^N (R_n^s - C_n^s)^2$ ；最后，选择二阶差分容量最小的方案作为最优波束等级组合方案，从而完成时隙分配与波束等级匹配的联合优化。

考虑到不同波束等级组合方案 s 下的时隙分配是互相独立的，因此可以并行执行动态时隙分配。最终，可将本文所提时隙分配和波束等级联合优化算法流程总结为图 6。

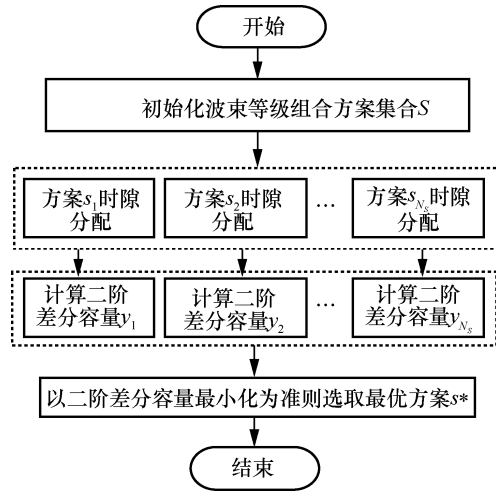


图 6 时隙分配和波束等级匹配联合优化算法流程

3 仿真结果与分析

3.1 场景和参数设置

本文采用的仿真工具为 MATLAB R2022a，并且在仿真时，假设 LEO 卫星工作于 Ka 频段，覆盖区域内包含 19 个小区。其余仿真参数如表 1 所示。

表 1	仿真参数
参数	值
工作频率 f /GHz	20
单波束最大发射功率 P_T /dBW	20
接收天线增益 G_n^R /dBi	42.1
跳波束周期 T_w	200
仿真总时长 T	$10T_w$
时隙持续时间 t /ms	90
卫星高度 H /km	550
卫星波束数 K /个	3
小区数目 N /个	19
小区边长 d /km	50
系统带宽 W /MHz	500
噪声温度 T^{noise} /K	300
雨衰系数	$\mu_s = 3.125, \sigma_s = 1.591$
乘性系数	$\beta = 2, \zeta = 2, \xi = 4$

3.2 仿真结果与分析

在进行仿真结果分析之前，本文先引入以下性能评估指标。

系统吞吐量。每轮实验传输的数据包总数。
波束重访时间。跳波束所分配的时隙之间的

间隔时间，重访时间过长将会影响用户终端的同步^[4]。

系统归一化平均冗余度。当方案 s 完成时隙分配后，通过计算该方案下实际业务处理量与实际业务需求量的差值并在归一化后取平均，即可得出该方案下的系统归一化平均冗余度，该指标可以直观反映出系统资源的利用情况，其表达式为

$$\bar{\eta} = \text{mean} \left(\left| \frac{C_1 - R_1}{R_1} \right|, \left| \frac{C_2 - R_2}{R_2} \right|, \dots, \left| \frac{C_N - R_N}{R_N} \right| \right) \quad (18)$$

其中， $\text{mean}(\cdot)$ 表示求元素的平均值， $|\cdot|$ 表示求元素的绝对值。

1) 混合波束等级方案优势分析

首先，为了体现混合波束等级带来的优势，本文考虑了以下 6 种场景。① 所有小区需求均较低；② 所有小区需求均较高，但都未超过单一等级波束最大处理能力；③ 大部分小区需求较低，剩余小区需求较高；④ 大部分小区需求较高，剩余小区需求较低；⑤ 中心小区需求较低，剩余小区需求较高；⑥ 中心小区需求较高，剩余小区需求较低。

然后，分别绘制了不同场景下混合波束等级方案、单一波束方案和单一等级 (L_1) 方案下的系统归一化平均冗余度曲线，如图 7 所示。

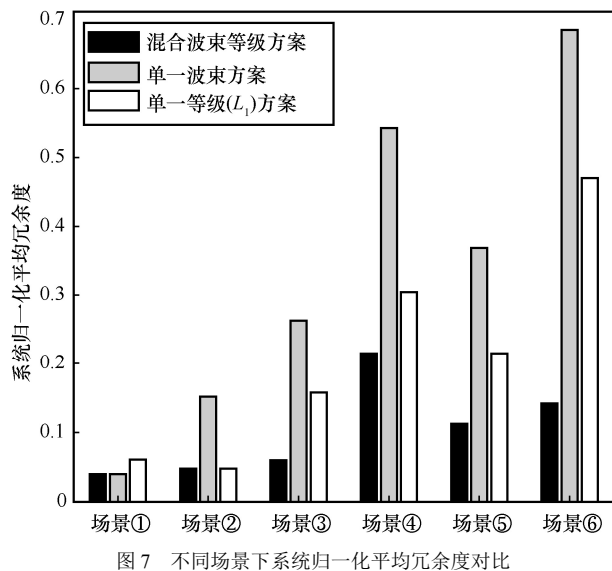


图 7 不同场景下系统归一化平均冗余度对比

由图 7 可知，在各种场景下，混合波束等级方案均保持了较低的系统平均冗余度。在场景①中，由于所有小区业务需求均较低，因此单一波束方案即可满足要求；在场景②中，由于所有小区需求均

较高且都未超过单一等级波束最大处理能力，因此场景②下采用单一等级 (L_1) 波束进行跳变即可达到最低系统归一化平均冗余度；在其余场景中，由于混合等级波束方案可以通过波束等级匹配来空出时隙以满足其余业务需求量较大的小区，因此当小区间业务差距较大时，混合波束等级方案依然可以表现出较低的系统归一化平均冗余度。

2) D-UCB 算法参数分析

对于 D-UCB 算法来说，折扣因子 γ 的选取将直接影响该算法的性能。当 $\gamma=1.0$ 时，D-UCB 算法将退化为传统 UCB 算法，并适用于奖励分布稳定已知的情况；当 $\gamma=0$ 时，D-UCB 算法则完全依赖于历史的奖励，只根据瞬时奖励做出决策。因此本文分别考虑了 $\gamma=0$ 、 $\gamma=0.5$ 以及 $\gamma=1.0$ 这 3 种典型情况下系统吞吐量随总业务需求量的变化情况，如图 8 所示。

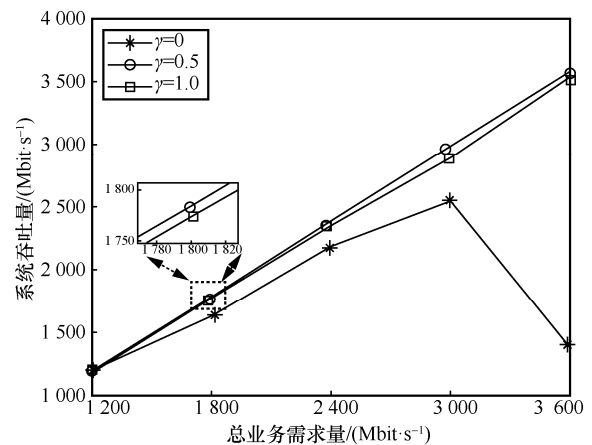


图 8 不同折扣因子下系统吞吐量变化对比

由图 8 可知，当 $\gamma=0$ 时，D-UCB 算法在总业务需求量较高时难以保证稳定的性能，且系统吞吐量始终难以满足业务需求量的要求；当 $\gamma=0.5$ 和 $\gamma=1.0$ 时，D-UCB 算法下的系统吞吐量均能保持稳定增长，且前者性能相对更高，这也反映出当奖励非稳定时，通过对历史奖励的选择性利用，可以让 D-UCB 算法保持较好的效果。

3) 不同算法性能对比分析

为了验证本文所提出的基于非稳态 MAB 模型的动态时隙分配算法的性能，本节考虑将该算法与平均时隙分配算法、贪婪时隙分配算法以及 ϵ 时隙分配算法进行对比，其中对比算法的具体介绍如下。

平均时隙分配算法。在跳波束周期中的每个时

隙依次选择一个小区进行时隙分配。

贪婪时隙分配算法。在跳波束周期中的每个时隙选择当前缓冲区数据包数量最大的小区进行时隙分配。

ε 时隙分配算法。在跳波束周期中的每个时隙按照 $1-\varepsilon$ 的概率对 D-UCB 算法进行时隙分配，否则依旧选择当前小区进行时隙分配。

最后，本文对比了不同算法下系统吞吐量情况，如图 9 所示。

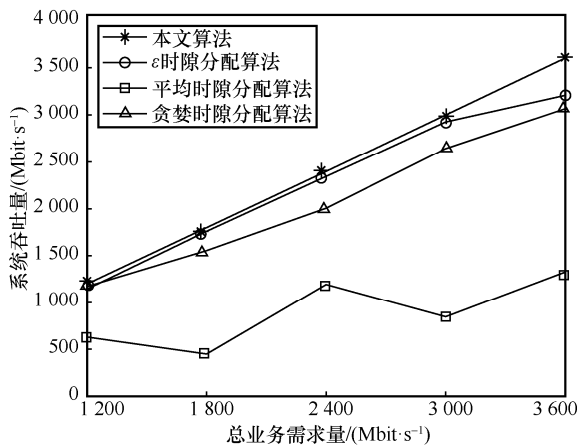


图 9 不同算法下系统吞吐量对比

由图 9 可知，随着总业务需求量的增加，平均时隙分配算法不能充分实现波束等级和时隙资源的灵活分配，因此其系统吞吐量始终较低；贪婪时隙分配算法只注重当前时隙的缓冲区数据情况，其决策从整个跳波束周期来看未必最优，因此其吞吐量相对较低； ε 时隙分配算法的性能在总业务需求量较低时比较接近本文算法，但由于其决策时仍然存在一定的随机性，因此该算法仍然存在一定的性能损失；相比这 3 种算法，本文算法的系统吞吐量随着总业务需求量的增加仍然能保持较稳定的增长，同时在总业务需求量较高时也能保持相对较好的性能。

4) 波束重访时间变化分析

波束重访时间也是衡量跳波束系统性能的一个重要指标。为了衡量本文算法的波束重访时间，本节进行了 50 次实验，并统计每次实验下的波束平均重访时间，最终结果如图 10 所示。

由图 10 可知，本文算法的波束平均重访时间始终保持在 300 ms 以内，且大部分情况下可以保持在 250 ms 左右，因此能够基本满足跳波束系统的同步性能要求^[5]。

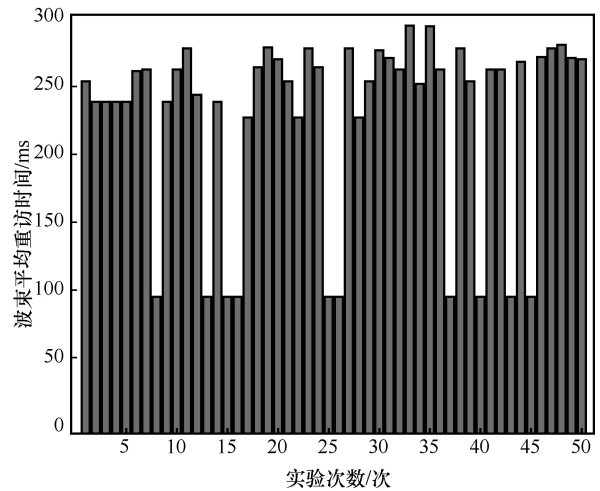


图 10 波束平均重访时间变化

4 结束语

本文研究了 LEO 跳波束卫星系统中的时隙分配和波束等级联合优化问题，从而在小区业务需求时变性和空间不均匀性的情况下实现卫星资源的灵活分配。相比于已有的跳波束算法，本文借助于非稳态 MAB 模型实现了在不依赖于环境信息以及较高计算资源下的动态时隙分配，且通过采用分级波束技术实现了波束覆盖范围与小区业务需求的匹配。仿真结果表明，相比单一波束方案和单一等级波束方案，本文所提混合波束等级方案的系统平均冗余度均不超过 20%；当总业务需求量从 1 200 Mbit/s 增长到 3 600 Mbit/s 时，本文所提动态时隙分配算法依然能保证较高的系统吞吐量，并将波束平均重访时间控制在 300 ms 以内，从而提高了 LEO 跳波束卫星系统的资源利用率。

参考文献：

- [1] 朱立东, 张勇, 贾高一. 卫星互联网路由技术现状及展望[J]. 通信学报, 2021, 42(8): 33-42.
ZHU L D, ZHANG Y, JIA G Y. Current status and future prospects of routing technologies for satellite Internet[J]. Journal on Communications, 2021, 42(8): 33-42.
- [2] SU Y T, LIU Y Q, ZHOU Y Q, et al. Broadband LEO satellite communications: architectures and key technologies[J]. IEEE Wireless Communications, 2019, 26(2): 55-61.
- [3] 张晨, 张更新, 王显煜. 基于跳波束的新一代高通量卫星通信系统设计[J]. 通信学报, 2020, 41(7): 59-72.
ZHANG C, ZHANG G X, WANG X Y. Design of next generation high throughput satellite communication system based on beam-hopping[J]. Journal on Communications, 2020, 41(7): 59-72.
- [4] LEI L, LAGUNAS E, YUAN Y X, et al. Beam illumination pattern design in satellite networks: learning and optimization for efficient

beam hopping[J]. IEEE Access, 2020, 8: 136655-136667.

- [5] 张晨, 彭明阳, 张更新. 基于联合优化的高通量卫星跳波束图案设计研究[J]. 南京邮电大学学报(自然科学版), 2021, 41(3): 1-8.
ZHANG C, PENG M Y, ZHANG G X. Beam hopping pattern method for high-throughput satellite based on joint optimization[J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science), 2021, 41(3): 1-8.
- [6] 王亚昕, 边东明, 胡婧, 等. 基于分簇的全带宽跳波束图案优化方法[J]. 计算机工程, 2020, 46(4): 169-176.
WANG Y X, BIAN D M, HU J, et al. Optimization method for full bandwidth beam hopping pattern based on clustering[J]. Computer Engineering, 2020, 46(4): 169-176.
- [7] TIAN F, HUANG L L, LIANG G, et al. An efficient resource allocation mechanism for beam-hopping based LEO satellite communication system[C]//Proceedings of 2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB). Piscataway: IEEE Press, 2020: 1-5.
- [8] LIU S J, HU X, WANG W D. Deep reinforcement learning based dynamic channel allocation algorithm in multibeam satellite systems[J]. IEEE Access, 2018, 6: 15733-15742.
- [9] BOUNEFOUF D, RISH I, AGGARWAL C. Survey on applications of multi-armed and contextual bandits[C]//Proceedings of 2020 IEEE Congress on Evolutionary Computation (CEC). Piscataway: IEEE Press, 2020: 1-8.
- [10] LI F, YU D X, YANG H, et al. Multi-armed-bandit-based spectrum scheduling algorithms in wireless networks: a survey[J]. IEEE Wireless Communications, 2020, 27(1): 24-30.
- [11] SEMENOV A, RYSZ M, PANDEY G, et al. Diversity in news recommendations using contextual bandits[J]. Expert Systems with Applications, 2022, 195: 116478.
- [12] GARIVIER A, MOULINES E. On upper-confidence bound policies for switching bandit problems[C]//International Conference on Algorithmic Learning Theory. Berlin: Springer, 2011: 174-188.
- [13] EVANS B G, THOMPSON P T, KYRGAZOS A. Irregular beam sizes and non-uniform bandwidth allocation in HTS[C]//Proceedings of the 31st AIAA International Communications Satellite Systems Conference. Reston: AIAA, 2013: 5617.
- [14] 镐梦婷, 张晨, 张更新. 基于相控阵的跳波束波位优化及资源分配算法[J]. 天地一体化信息网络, 2022, 3(1): 18-26.
GAO M T, ZHANG C, ZHANG G X. Beam hopping beam optimization and resource allocation algorithm based on phased array[J]. Space-Integrated-Ground Information Networks, 2022, 3(1): 18-26.
- [15] JOHN D K, RONALD J M. Antennas: for all applications[M]. 3rd ed. New York: McGraw-Hill, 2001.

[作者简介]



林敏 (1972-), 男, 浙江台州人, 博士, 南京邮电大学教授、博士生导师, 主要研究方向为无线通信系统、智能信号处理、天线新技术等。



阚鹏程 (1999-), 男, 江苏扬州人, 南京邮电大学硕士生, 主要研究方向为卫星通信、无线通信、人工智能。



赵柏 (1995-), 男, 江苏泗阳人, 南京邮电大学博士生, 主要研究方向为无线通信中的智能信号处理技术、人工智能、混合多址接入技术等。



程铭 (1991-), 男, 江苏常州人, 博士, 南京邮电大学讲师, 主要研究方向为星地融合网络、无线通信中的性能评估、人工智能等。



杨绿溪 (1964-), 男, 安徽桐城人, 博士, 东南大学教授、博士生导师, 主要研究方向为无线通信中的信号处理、MIMO通信、智能无线通信和统计信号处理等。