

面向巨型星座系统的多地面站协同测控技术

刘 阳¹, 周 笛¹, 盛 敏¹, 李建东¹, 郝时光², 郑晓天²

(1. 西安电子科技大学综合业务网理论及关键技术国家重点实验室, 陕西 西安 710071)

(2. 中国空间技术研究院通信与导航卫星总体部, 北京 100094)

摘 要: 测控技术是保障星座系统高效运维和管理的关键技术。近年来, 随着星座规模的不断扩大, 逐步形成了巨型星座系统, 使得对星座的测控需求呈现爆发式的增长, 从而对星座系统测控任务的完成量提出了新的要求。首先分析巨型星座系统测控任务约束和地面测控站设备约束并给出问题建模; 其次提出一种地面站Agent基于学习规划片段的交互方式, 通过引入约束惩罚算子和多地面站联合惩罚算子设计优化的目标函数; 最后, 提出一种多地面站Agent强化学习算法以求解多地面站协同任务分配策略。仿真结果显示, 任务规模较大时, 在文中提到的不同场景下该方法较传统算法有12%~20%的增益。

关键词: 巨型星座; 测控任务规划; 多地面站协同; 多Agent强化学习

中图分类号: TP393

文献标志码: A

doi: 10.11959/j.issn.2096-8930.2023001

Multi-Ground Station Collaborative Measurement and Control Technology for Giant Constellation System

LIU Yang¹, ZHOU Di¹, SHENG Min¹, LI Jiandong¹, HAO Shiguang², ZHENG Xiaotian²

1. The State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China

2. Communications and Navigation Satellite General Department, China Academy of Space Technology, Beijing 100094, China

Abstract: Measurement and control technology is the key technology to ensure the efficient operation, maintenance and management of the constellation system. In recent years, with the continuous expansion of the constellation scale, mega-constellation system has gradually formed, which makes the demand for constellation measurement and control show an explosive growth, which puts forward new requirements for the completion of the constellation system measurement and control tasks. Firstly, the constraints of the mega-constellation system measurement and control tasks and the equipment constraints of the ground measurement and control station were analyzed, and the problem modeling was given; Secondly, an interaction method of the ground station agent based on the learning planning segment was proposed, by introduced the constraint penalty operator and the multi-ground station joint the penalty operator was designed to optimized the objective function. Finally, a multi-ground station Agent reinforcement learning algorithm was proposed to solved the multi-ground station cooperative task assignment strategy. Simulation experiments showed that when the task scale was large, the method had a gain of 12%~20% compared with the traditional algorithm in the different scenarios mentioned.

Keywords: mega constellation, measurement and control task planning, multi-ground station coordination, multi-Agent reinforcement learning

收稿日期: 2022-11-15; 修回日期: 2023-02-27

通信作者: 周笛, zhoudi@xidian.edu.cn

基金项目: 国家自然科学基金资助项目 (No.U19B2025, No. 62121001, No. 62001347); 陕西省重点研发计划 (No. 2022ZDLGY05-02)

Foundation Items: The National Natural Science Foundation of China (No.U19B2025, No. 62121001, No. 62001347), Key Research and Development Program of Shaanxi (No. 2022ZDLGY05-02)

0 引言

构建空间互联网^[1-2]和空间物联网的理念^[3-4]，以及将卫星网络融入6G系统的愿景^[5-7]，再次激发了卫星网络在未来无线网络中的研究热潮^[8]。不同于几十年前的网络配置^[9]（如规模和结构），低轨道卫星星座如Starlink和OneWeb通过逐渐趋向密集和规模化来提升网络覆盖率和全球宽带提供能力^[2]。最终，LEO卫星星座将扩展为巨型星座（MSC）^[1]。这对利用现有有限个地面测控站来保障巨型星座的稳健运行带来了严峻的挑战^[8]。测控技术是保障星座系统高效运维和管理的关键技术，作为航天工程中不可或缺的一部分，航天测控承担着航天器飞行轨道与飞行姿态的调整、星上载荷的控制以及各模块工作状态的监控工作^[10]，是保障航天器正常运作并完成既定任务的重要基础，此外对航天器实时的监控能够对突发事件进行快速处理甚至预警^[11]。

由于卫星产业的进步，各类卫星的发展都极其迅速^[12]，在轨航天器的数量、类型及其功能愈加丰富，对地面设备有更多的要求，不同的地面站部署着不同功能的设备，这都增加了测控任务的复杂程度。此外由于当今国际局势的复杂性，国外布站的成本较高，提高地面测控资源的利用率十分必要。除了设备因素，随着卫星星座规模不断扩大，空间信息网络的任务数量急剧上升，复杂的任务需求和资源调控对测控系统的全局掌控能力也提出了更高的要求^[13]。综上所述，伴随星座系统规模的扩大、任务和设备的复杂化，测控系统的管理人员必须采用及时高效的规划算法整合资源做出规划。

目前在巨型星座系统常规测控方面的研究如下。

采用经典的数学模型来解决，参考文献[14]设计了一种两阶段无冲突测控请求任务数最大化的方法。参考文献[15]提出了一种非线性泛函模型，并在忽略释放阶段对任务规划影响的假设条件下，将其简化为带约束的0-1整数规划模型。

采用图论的方法来解决，参考文献[16]将弧段和约束构造为图，以最短工作时长作为目标函数，根据工作时段将资源调度问题转化为子集优化问题。参考文献[17]采用了Petri网络，通过对需求及设备类型、实际调度规则等系统要素的分析，建立了多星一站系统的运行模型。

采用机器学习等算法，参考文献[18]建立了以最大化调度方案总权重为目标的约束满足模型，对单天线地面站系统任务调度问题进行动态规划。参考文献[19]将测控资源调度系统划分为任务Agent、网管Agent和资源Agent三大类，建立了航天测控任务规划的多智能体系统（Multi-Agent System, MAS）模型。

尽管目前已经有针对常规测控任务规划问题的广泛

研究，但大多数研究建模的时候仅仅当作时间窗规划任务，没有将现实系统中复杂约束和任务优先级等特点加入模型中；此外，对于中小规模测控调度问题，上述算法能够得到最优或近似最优的解。但是在巨型星座系统的实际应用中，测控需求和测控设备众多，问题规模庞大，由于求解时间和空间的限制，上述算法难以得到最优解或近似最优解。因此，上述算法并不能直接应用于未来巨型星座系统场景，需要进一步探索新的方法。本文在分析测控任务特点和卫星约束的基础上，将测控任务规划问题刻画成一个分布式约束优化问题，给出多地面站协同任务规划的数学模型，避免在巨型星座网络中必须要由一个或者几个节点获取并统一规划整个网络中其他大量节点测控资源的情况出现，缓解传统集中式测控规划造成的计算压力。其次提出一种地面站Agent基于学习规划片段的交互方式，通过引入约束惩罚算子和多地面站联合惩罚算子设计了优化的目标函数。在此基础上提出一种多地面站Agent强化学习算法以求解多地面站联合协同任务分配问题。

1 系统模型和问题建模

1.1 网络模型

图1为本文针对巨型星座系统的网络场景。该系统主要由地面测控网络和低轨卫星星座组成。卫星周期性地按照轨道运行，在可见弧段内完成测控任务。地面测控网络主要包括用户端、测控中心和地面站。用户通过用户端设备将需求提交给测控中心或地面站，并通过测控中心或者地面站获取所需数据。测控中心负责将用户提交的需求翻译成可以被系统理解的任务请求列表，再将请求列表通过地面通信网络下发给各地的地面站。具有自主规划能力的地面站通过与低轨卫星在可见范围内进行通信完成指令上注和数据回传等任务，还能通过地面网络分享自己的信息。

1.2 测控任务规划流程

在实际的测控任务规划模型中，本文考虑众多分布在全国不同经纬度地区的测控地面站、依赖测控通信网连接地面测控站的测控中心以及分布于不同轨道的低轨卫星构成的卫星测控场景。在卫星周期性的运动过程中，当与地面站的视线（Line of Sight, LoS）存在时认为产生了可见时间窗，在可见时间窗之内地面站才有可能与卫星通信，测控任务流程如图2所示。

在一定的时间段内，位于测控中心的操作员汇总来自各地卫星用户的服务需求，并希望建立测控中心与卫星的联系，从而指示卫星执行相应的动作，但是这一

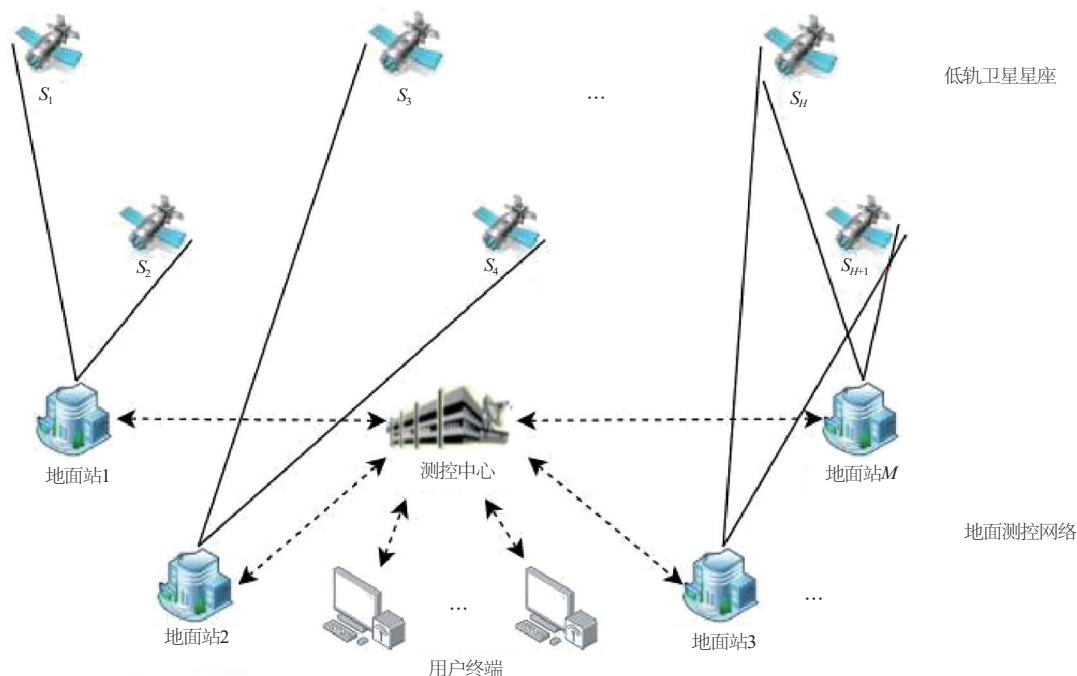


图1 网络场景

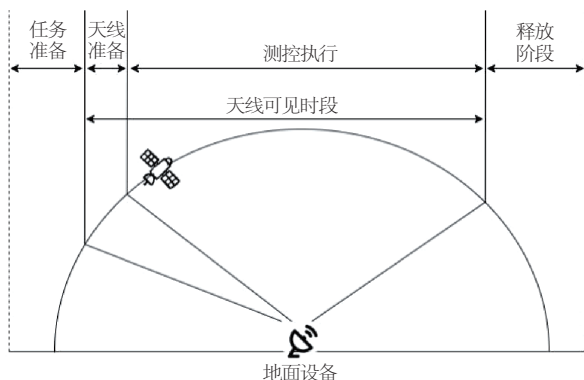


图2 测控任务流程

过程必须借助地面站网络才能进行，测控中心无法直接对卫星进行任务指令上注。基于场景的动态性以及任务需求，操作员将任务需求翻译为可被卫星执行的操作计划。此外，卫星经过地面站上空之前测控中心会得到相关的卫星预报，说明地面站对系统中卫星的可见弧段资源。根据已知的操作计划和卫星预报，操作员将产生一系列包含时间窗约束的请求集合。测控规划的目标就是根据请求集合，产生一个最终决定的规划列表。

1.3 任务需求与约束模型

为了解决多星任务规划问题，需要明确任务属性、卫星和地面设备约束^[20]。本文研究的卫星指令上注阶

段，即测控任务，包含的主要要素如下。

(1) 任务指定的卫星。测控任务是由地面测控中心通过地面设备向指定卫星发送指令的过程，卫星只有接收到了来自地面设备的指令才有可能在合适的时间执行该指令，从而满足用户需求。

(2) 任务收益。测控任务数量众多，且不同的任务有不同的紧急程度，为了刻画这种紧急程度，常采用优先级这一参数。同时可以将任务的优先级视为完成该任务所能得到的收益。

(3) 任务完成的时间窗。为确保卫星及时执行任务，每个任务的执行指令都要在指定的时间区间内完成指令上注。由于地球自转以及卫星轨道等因素，卫星每次掠过地面站的角度和时间不同。测控任务的执行时间晚于最早开始时间，早于最晚结束时间并且不能短于最短执行时长。对于固定时间窗任务，则只能在指定的时间段内完成指令上注。

卫星、地面站是执行测控任务的两个主体，二者具有不同的属性和约束条件。对它们的分析如下。

(1) 卫星轨道。卫星发射后将进入固定轨道。如图3所示，对卫星轨道一般用图中所示的6个参数进行刻画。对有些测控任务的规划，如果在指定的时间内卫星与地面站不可见或者可见的时间长度不够，将导致规划失败。

(2) 卫星姿态。为了确定卫星位置，一般借助地球坐标系 $Ox_oY_oZ_o$ 和自身坐标系 $X_sY_sZ_s$ 。如果卫星与地面站

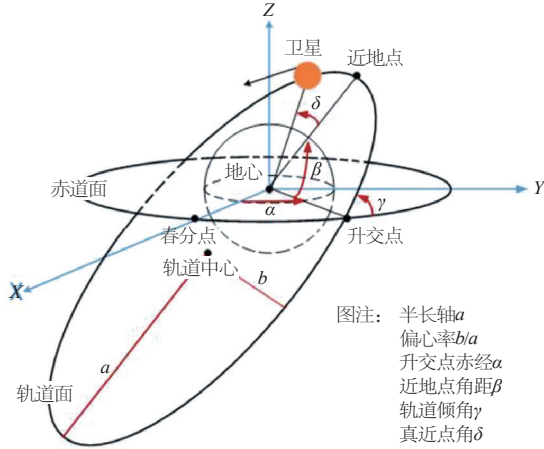


图3 卫星轨道示意

进行通信时卫星天线和地面设备的天线不在同一条直线上，那么需要调整卫星姿态，使地面设备对准天线。

(3) 地面设备开机准备。每次执行任务之前，卫星和地面设备都需要做一定的准备工作。卫星需要调整姿态使天线对准地面设备的天线；地面设备在和指定卫星建立通信之前，需要提前5 min准备，并且在传输过后，要有1 min的时间释放。

(4) 俯仰角约束。在可见弧段之间，地面设备不断追踪卫星位置。在此期间卫星不断移动，地面设备的仰角也随之变化。在实际工作中，由于电磁波直线传播、地球曲率、卫星轨道等因素，一般认为天线仰角在3°以下时，地面设备与卫星不可见。

1.4 多地面站协同规划问题建模

由于巨型星座系统日常运维以及各地卫星用户的任务需求，在规划时间内将产生大量测控请求；相对的每个具有自主规划能力的地面站的测控资源有限，测控资源的分配无法仅仅依靠一个或者几个中心节点统一获取和规划。地面站之间需要通过地面网络进行交互，合作完成规划任务，将测控任务目标合理地分配给各个地面站，使得总体收益最大。在协同规划过程中有如下变量。

(1) 任务规划区间 $\omega_{\text{schedule}}=[t_s, t_e]$ ， t_s 和 t_e 分别表示任务规划的起始时间和结束时间。

(2) 测控网络系统中存在 N_G 个测控地面站Agent，表示为 $\text{GS}=\{\text{gs}_1, \text{gs}_2, \dots, \text{gs}_{N_G}\}$ 。 $\forall \text{gs}_k \in \text{GS}$ 表示为 $\text{gs}_k = \langle R_{\text{st}}^k \rangle$ ， R_{st}^k 表示在任务规划区间内地面站 gs_k 可提供的系统中各颗卫星的可见时间窗资源。这里各个地面站对于同一卫星的可见时间窗资源需求向量不相同。

(3) 系统中存在 N_s 颗待测卫星 $\text{SAT}=\{\text{sat}_1, \text{sat}_2, \dots, \text{sat}_i, \dots, \text{sat}_{N_s}\}$ ，巨型星座系统中 N_s 为几百或者几千颗。

(4) 规划时间内系统中存在测控任务 $T=\{t_1, t_2, \dots, t_j, \dots, t_{N_t}\}$ 可表示为 $t_j = \langle \text{sat}_j, u_j, A_j \rangle$ ，分别表示该测控任务对应的卫星 sat_j ，完成指派给卫星的测控任务所能获取的效用以及完成任务所占用的可见时间窗资源。

基于以上的描述，多地面站协同任务规划的实质是地面站之间进行自主的互动协商，以确定每个地面站的测控任务集合 ST_k ，其中 $\text{ST}=\bigcup_k \text{ST}_k$ ， $\text{ST}_k \in T$ 。各地面站的测控任务分配需要在满足约束条件的前提下，实现总体收益最大化。总体收益通过收益函数 $V(T)$ 给出。

$$V(T) = \sum_{k \in \text{GS}} \sum_{i \in \text{ST}_k} u_i \cdot x_i^{(k)} \quad (1)$$

$$x_i(k) = \begin{cases} 1, & \text{设备gs完成对任务} t_i \text{的测控} \\ 0, & \text{其他} \end{cases} \quad (2)$$

优化的目标子集为 T^* ，即

$$V(T^*) = \text{MAX}(V(T)) \quad (3)$$

有如下约束： $\forall t_i, t_j \in T$ 且 $i \neq j$

$$r_{\text{st}}^i(k) \cap r_{\text{st}}^j(k) = \emptyset \quad (4)$$

其中， $r_{\text{st}}^i(k) = [s_i^k, e_i^k]$ ， $r_{\text{st}}^i(k) \in R_{\text{st}}^k$ 表示测控任务 t_i 占用测控设备 gs_k 的时间窗资源。 s_i^k ， e_i^k 分别表示任务的开始时间和结束时间。式(4)保证了同一个地面站分配的不同测控任务所需的时间窗资源不冲突。

2 基于多智能体强化学习的规划算法

2.1 效用值增益函数

首先给出多地面站强化学习的相关定义。

定义1 地面站Agent的状态 S_t^k 。重点关注地面站的两个维度：当前的剩余测控资源和当前的学习规划片段。

也即用二维元组 $\langle \text{CR}_t^k, \text{CT}_t^k \rangle$ 表示 t 时刻地面站Agent所处的学习状态。其中 CR_t^k 为地面站Agent当前还可提供的时间窗资源； $\text{CT}_t^k \in T$ 表示学习规划片段，具体内容为地面站 t 时刻所规划的测控任务目标集。地面站Agent所有可能处于的状态组成一个集合，称之为状态集，即

$$S_k = \{S_t^k \mid t=1, 2, \dots\}$$

定义2 地面站的累计效益 u_t^k 是指 gs_k 在状态 S_t^k 下已经规划的所有任务的效益之和，计算方法为

$$u_t^k = \sum_{t_i \in \text{CT}_t^k} u_i$$

定义3 地面站 gs_k 的动作集合 $\text{ACT}_k = \{a_t^k \mid a_t^k \in P(T)\}$ ，其中 $P(T)$ 表示目标集合 T 的幂集。动作集合仅仅包含幂集中可被 gs_k 执行的测控任务。

定义4 地面站的更新效益 ω_t^k 为地面站执行 a_t^k 中的测控任务时所能得到的任务效益的总和，表示为

$$\omega_t^k = \sum_{t_i \in a_t^k} u_i$$

地面站 gs_k 在状态 s_t^k 根据策略选择执行的动作 $a_t^k \in \text{ACT}_k$ ，将导致该地面站Agent转移到下一个状态 s_{t+1}^k ，它与 s_t^k 有如下关系

$$s_{t+1}^k = \langle \text{CR}_{t+1}^k, \text{CT}_{t+1}^k \rangle \quad (5)$$

$$\text{CT}_{t+1}^k = a_t^k \quad (6)$$

现在先考虑单地面站条件下的学习策略，记 $f_k(s_t^k, a_t^k, s_{t+1}^k)$ 为地面站Agent在状态 s_t^k 下采取行动 a_t^k 转移到下一状态 s_{t+1}^k 的效用值增益，根据式(7)计算

$$f_k(s_t^k, a_t^k, s_{t+1}^k) = \omega_t^k - \mu_t^k - \eta_t^k \quad (7)$$

这里 η_t^k 为约束惩罚算子。它的计算方法是取值为 a_t^k 规划的任务中违反式(4)约束的任务的效用值乘以系数 λ 。如果采取的动作中的所有任务没有违反任何约束，那么 η_t^k 值为零，否则大于零。学习优化的目标就是各个地面站独立地将测控资源尽可能多地分配给测控需求且尽可能少地违反时间窗约束。

考虑多站协同规划中的地面站Agent学习策略，每个地面站仍可以采用式(7)的效用值增益作为回报构建强化学习算法和策略、参数更新的机制。这种方法可以避免常规算法“建模难”的问题，但是每一次迭代都需要在整个动作空间中搜索。当系统规模和任务规模过大时，算法收敛得十分缓慢。此外，由于各地面站独立搜寻使自身规划任务收益最大的方案，有可能导致在整个系统中多个地面站对多数相同测控目标的重复规划，这就造成测控资源的浪费。

针对以上两点，本文基于TAN^[22]交互学习的思想，设计地面站Agent的基于学习规划片段的交互方式。地面站Agent在学习的每一步中彼此交互当前的学习规划片段，每个Agent都能根据其他Agent的学习规划片段改进自身的规划选择，从而消解时间窗冲突；此外，结合本文引入的联合惩罚因子 δ_t^k ，对统一测控任务重复规划的情况施加惩罚。联合效用值增益如下

$$f_k(s_t^k, a_t^k, s_{t+1}^k) = \omega_t^k - \mu_t^k - \eta_t^k - \delta_t^k \quad (8)$$

联合惩罚因子的计算流程如下。

算法1 联合惩罚因子计算

输入: $\text{CT}_{t+1}^k, \text{Epd}'_{\text{other}}$

输出: δ_t^k 联合惩罚因子

```

1:  $\delta_t^k = 0$ 
2: for each  $t_i \in \text{CT}_{t+1}^k$ 
3:   count = 0
4:   do for each  $\text{CT}_i^j \in \text{Epd}'_{\text{other}}$ 
5:     for each  $t_l \in \text{CT}_i^j$ 
6:       do if  $t_i \in t_l$ 

```

```

7:       count = count + 1

```

```

8:        $\delta_t^k = \delta_t^k + \left( \frac{\text{count}}{\text{count} + 1} \right) \cdot u_i$ 

```

```

9:   return  $\delta_t^k$ 

```

算法1中， $\text{Epd}'_{\text{other}} = \{\text{CT}_i^j | gs_j \in \text{GS}, j \neq k\}$ ，表示其他地面站当前的学习规划片段的集合。多Agent强化学习的多地面站协同任务规划中，地面站Agent需要提供自己的学习规划片段，同时访问其他地面站Agent的学习规划片段，并依据这些信息选择规划动作。地面站Agent通过图4所示的方式进行规划片段的交互。

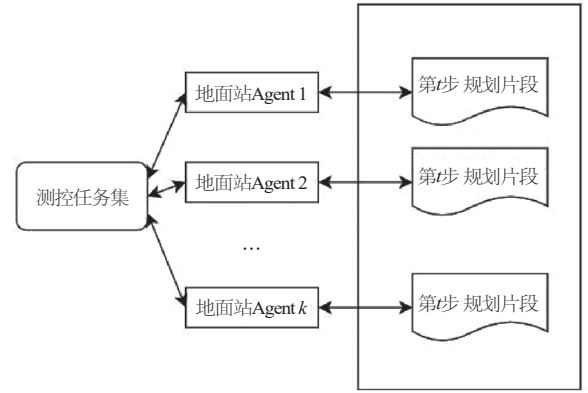


图4 学习规划片段的交互

2.2 基于“输快赢慢”以及策略爬山的多站协调规划算法

在传统的多智能体Q学习算法的基础上分析劣势并提出多地面站智能体强化学习算法（Multiple Ground-Station Agent Reinforcement Learning Algorithm, MUGARLA），详细介绍算法原理和实现步骤。

在多站协同规划问题中，本文假设各个地面站间是合作的关系。当测控目标未被重复访问时，提高任意地面站规划收益都能提升整个系统的总收益。但是每个地面站的能力有限且互异。因此需要每个地面站Agent不断地与环境（包括其他地面站）交互并进行自身的学习，达到总体上最优的联合规划结果。

本文采取多智能体Q学习（Multiple-Agent Q-Learning, MAQL）对上述问题进行求解。学习过程中，每个Agent每一步在有限可行动作集合中依据策略选择动作，执行该动作后原本的状态发生改变，通过与环境的交互得出该状态-动作对的回报，从而不断学习，使得值函数尽可能地大，达到较优的状态。在巨型星座测控系统中，系统的地面站、测控目标卫星以及测控任务规模都较大，由于每个Agent的策略不仅需要考虑自身状态，而且需要参考其他Agent的信息，使得每个Agent

的状态空间非常大；此外动作空间也面临类似的“维数灾难”情况。

鉴于此，本文通过一种基于学习规划片段的交互方式改进上述的MAQL，即WoLF-PHC (Win or Learn Fast-Policy HillClimbing) 算法。算法在各地面站Agent的交互协商过程中，并不直接利用其他Agent的所有信息，大大缩小了每个Agent的动作空间集和状态集。彼此之间的交互仅仅依赖于学习规划片段，通过多站联合惩罚因子表征冲突大小，提高了多地面站的协作能力，改善了学习优化性能。另外，与Nash-Q等算法相比，WoLF-PHC只需要保存相对较少的信息，可以降低多Agent问题求解时的空间复杂度。

该算法中，地面站Agent的策略仅受自身状态影响，记为 π_k 。初始化时，所有状态下任意行动的选择概率都相等，为对应Agent动作空间的平均。即对 $\forall s'_k, a'_k, \pi_k = \frac{1}{|\text{ACT}_k|}$ 在第 t 个学习阶段， gs_k 更新决策按照以下原则

$$\pi_k(s'_k, a'_k) = \pi_k(s'_k, a'_k) + \begin{cases} \varepsilon_k, a'_k = \arg \max_{a' \in \text{ACT}_k} Q_k(s'_k, a') \\ -\varepsilon_k, \text{其他} \end{cases} \quad (9)$$

其中，学习增量 ε_k 随着学习的进行逐渐下降。 $Q_k(s'_k, a')$ 表示 gs_k 的状态-行动对 (s'_k, a') 的 Q 值。在WoLF-PHC算法中，将依照“输快赢慢”的原则进行策略更新。结合 Q 值表和平均学习期望来判断每一步学习的“输”或“赢”，并选择学习增量 ε_k 取值如下

$$\varepsilon_k = \begin{cases} \varepsilon_{\text{lose}}, & \text{如果为loss} \\ \varepsilon_{\text{win}}, & \text{如果为win} \end{cases} \quad (10)$$

判断依据是如果

$$\sum_{a'_k \in \text{ACT}_k} \pi_k(s'_k, a'_k) Q_k(s'_k, a'_k) > \sum_{a'_k \in \text{ACT}_k} \bar{\pi}_k(s'_k, a'_k) Q_k(s'_k, a'_k) \quad (11)$$

则判定为“赢”，否则为“输”。

一般情况下 $\varepsilon_{\text{lose}}$ 是 ε_{win} 的几倍。在本文的仿真实验中， $\varepsilon_{\text{lose}}$ 是 ε_{win} 的4倍。其中 $\bar{\pi}_k(s'_k, a'_k)$ 是平均策略，其更新有以下公式

$$\bar{\pi}_k(s'_k, a'_k) = \bar{\pi}_k(s'_k, a'_k) + \frac{\pi_k(s'_k, a'_k) - \bar{\pi}_k(s'_k, a'_k)}{c_k(s'_k)} \quad (12)$$

式中， $c_k(s'_k)$ 是地面站 gs_k 到达状态 s'_k 的次数。

通过前面的分析，得到地面站 gs_k 的 Q 学习公式如下

$$Q_k(s'_k, a'_k) = Q_k(s'_k, a'_k) + \alpha [f_k(s'_k, a'_k, s_{t+1}^k) + \gamma \max_{a' \in \text{ACT}_k} Q_k(s_{t+1}^k, a')] \quad (13)$$

其中， $\alpha > 0$ 为学习率， $\gamma (0 \leq \gamma < 1)$ 表示折扣因子。

算法的整体框架如下。

算法2 多地面站智能体强化学习算法

输入：系统中的地面站集合GS，测控目标集合 T

输出：每个地面站的策略 π_k ，学习规划片段 Epd_k

1：根据系统内配置（地面站和测控目标卫星）计算每个地面站的可达访问任务集合。初始化每个地面站保存的 Q_k 表。初始化每个地面站的规划任务列表。随机初始化每个地面站的任务选择策略 π_k ，令 $t=0$ 。

2：对学习的每一步，令 $k=0$ 。

3： gs_k 获取其他地面站的学习规划片段 $\text{Epd}'_{\text{other}}$ 。

4： gs_k 根据自己的当前策略表 π_k 在 s_t^k 下选取规划动作 a_t^k 。

5：计算 s_{t+1}^k ；保存观测样本数据 $\langle s_t^k, a_t^k, s_{t+1}^k, \omega_t^k \rangle$ 。

6：根据式（8）计算联合增益函数 $f_k(s_t^k, a_t^k, s_{t+1}^k)$ ，根据式（13）更新 $Q_k(s'_k, a'_k)$ 。

7：根据式（9）更新 π_k 在当前状态下的动作选择概率。

8：更新 gs_k 的学习规划片段 $\text{Epd}'_{\text{other}}$ 。

9： $k=k+1$ ，若 $k=|GS|+1$ ，继续下一步，否则返回步骤3。

10：若满足学习终止条件，则学习结束，否则 $t=t+1$ ，返回步骤2。

3 仿真验证

3.1 仿真参数

本文仿真过程中，设置强化学习相关参数见表1、表2。

仿真中所选取的卫星取自表3中所显示的巨型星座构型，星座中的卫星位于覆盖全球的不同的轨道面上。

本文实验通过STK与MATLAB联调模拟仿真场景中的星座和地面站并且输出可见时间窗数据。用Python产生测控需求并具体实现各个算法的仿真，验证本文所提算法的性能。仿真分为两个部分：第一部分给出地面站间是否进行合作（学习规划片段交互）的性能对比；第二部分给出本文提出的算法与贪婪算法、随机算法在测控任务不同任务及场景规模情况下的性能比较。

3.2 仿真结果

第一部分，模拟卫星任务为在24 h内返回卫星运行所需状态数据用作网络维护。与此对应的测控任务为24 h内地面站尽量对系统中所有卫星进行至少一次指令上注。即任务目标为系统中的1 134颗卫星，测控任务规模为1 134颗，系统24h内的弧段预报总数在40万次以上。

在总计20万步的学习过程中采样100个状态点，统计已规划的任务的总收益，仿真结果如图5所示。其中

表1 学习相关参数

参数	值
学习率 α	0.1
折扣因子 γ	0.9
ϵ_{loss}	0.002 5
ϵ_{win}	0.01
迭代总步数	20万步

表2 场景数据集

地面站个数(128个)	场景1	场景2	场景3	场景4
卫星数/颗	189	378	756	1134
卫星数(1134颗)	场景5	场景6	场景7	场景8
地面站个数/个	47	83	128	162

表3 巨型星座构型

星座名称	轨道数/个	每轨星数/颗	总星数/颗	轨道倾角/ $^{\circ}$	轨道高度/km
LEO-1	18	21	378	85	1 345
LEO-2	36	21	756	89	1 100

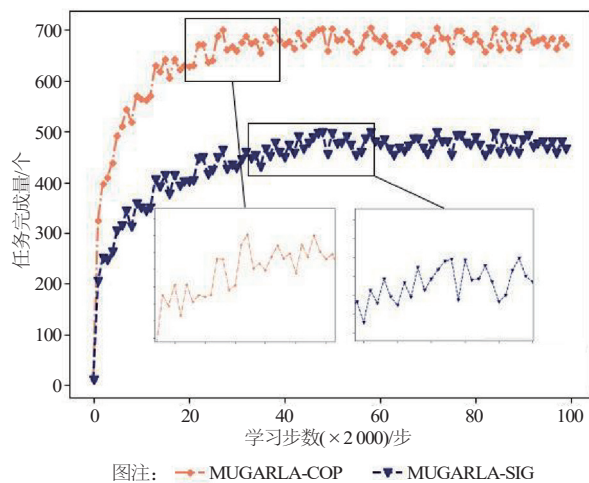


图5 MUGARLA-COP和MUGARLA-SIG学习过程

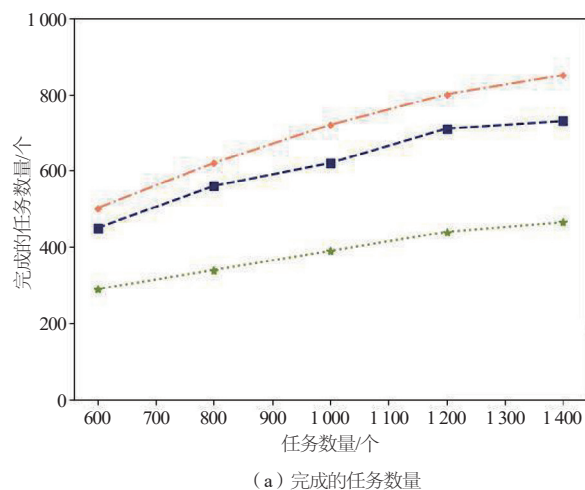
MUGARLA-SIG表示了地面站Agent按照式(7)进行独立学习的结果, MUGARLA-COP表示了地面站Agent采用式(8)进行交互学习的结果。从结果来看, 两种方法都能得到收敛的结果, 然而MUGARLA-COP在学习过程中地面站之间彼此交流当前的学习规划片段。这一方面扩展了Agent“环境”的理解范围, 丰富了学习知识, 因此算法收敛较之独立学习要快。在图5中表现为MUGARLA-COP的学习图线坡度较独立学习要陡, 且峰值的出现先于MUGARLA-SIG的学习图线, 图线更早地在峰值上下波动; 另一方面在交互过程中能够更为直接地检测到规划任务之间的重复与冲突, 有利于算法跳

出局部最优, 提高了可见窗资源的利用率, 所以最终获得了更高的收益。

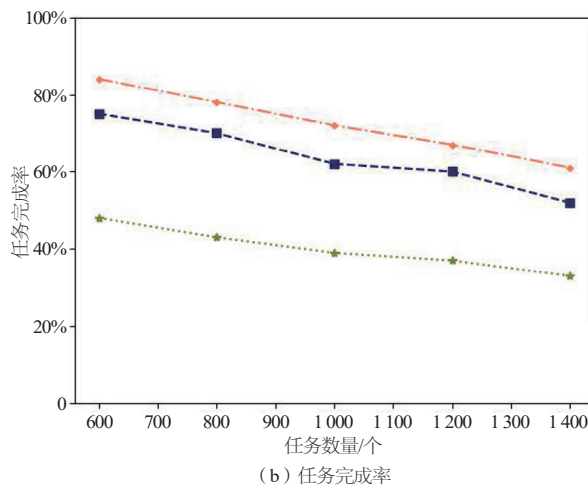
为了证明算法的有效性, 第二部分在不同的测控任务规模下, 将MUGARLA-COP与贪婪算法和随机算法分别在不同卫星规模以及任务数量的情形下进行比较。其中贪婪算法的实现是基于时间窗结束时间的贪婪算法; 随机算法的实现是每个地面站在可选任务列表中随机选取任务集合。

假设系统中的卫星绕轨道运动发生2次升圈次, 2次降圈次, 本文仿真在卫星的升圈或降圈过程中随机产生需求。对于测控需求规定了任务的时间窗(最早开始时间和最晚结束时间)以及最短持续时间, 意味着地面站对卫星的可见弧段在测控需求的时间窗内时, 该测控任务才能被执行。图6展示了3种算法在不同需求规模情况下的任务完成情况。

从图6可以看出, 在小规模的任务条件下, 任务冲



(a) 完成的任务数量



(b) 任务完成率

图注: —●— MUGARLA-COP —■— 贪婪算法 —▲— 随机算法

图6 3种算法在不同需求规模情况下的任务完成情况

突程度较低，MUGARLA-COP和贪婪算法的任务完成情况相差较小。随着任务规模和冲突程度的增加，MUGARLA-COP较贪婪算法取得了更好的规划结果。随机算法各地面站在选择可执行任务集合时完全采取随机的方式，导致所规划的任务间出现大量的冲突和重复，从而任务完成情况不佳。

MUGARLA-COP引入了学习规划片段，所有的地面站在每一步的学习过程中都能够得知上一步系统中其他地面站的任务规划情况；由于引入了联合惩罚算子，能对系统中重复规划的任务进行消解，因此取得了较好的规划结果。贪婪算法在任务规模较小时能得到整体不错的解，但是随着规模的变大，整体规划性能下降。从图6中能看到MUGARLA-COP在任务需求过大时的任务完成率也在下降，这主要是由动作集合的选择策略导致的。在任务规模较小的时候，地面站根据自身可见资源和测控需求所要求的时间窗情况所能规划的可行任务数有限。在这种情况下，通过将可行任务数的幂集作为可行任务集合，在足够

的学习过程中就可以遍历所有情况从而达到最优的规划结果。在地面站可行任务数增长至一定规模时，为了避免指数爆炸，一方面要在预处理中删除包含任务数量较少的集合，保证测控资源使用率；另一方面通过优先级排序，将优先级较高的需求更多地包含在可行任务集合中，并控制可选任务集合的个数至不低的水平。这样做会人为带来更多重复规划的可能性，但是能够保证优先级高的任务得到优先规划，也能进一步降低动作空间的规模，保证算法的正常运行。

图7展示场景1~场景4下的4种算法性能表现。4种场景下的卫星数分别为189颗、378颗、756颗、1 134颗（场景2+场景3中的卫星）。地面站个数为128个。系统中的卫星在24 h内各随机产生0~2个测控需求。

从上面的数据可知，整体上MUGARLA-COP的算法有效性优于其他算法，贪婪算法次之，而随机算法的性能较差。其中卫星规模以及任务需求规模较低时，MUGARLA-COP和贪婪算法都能保证较高的任务完成

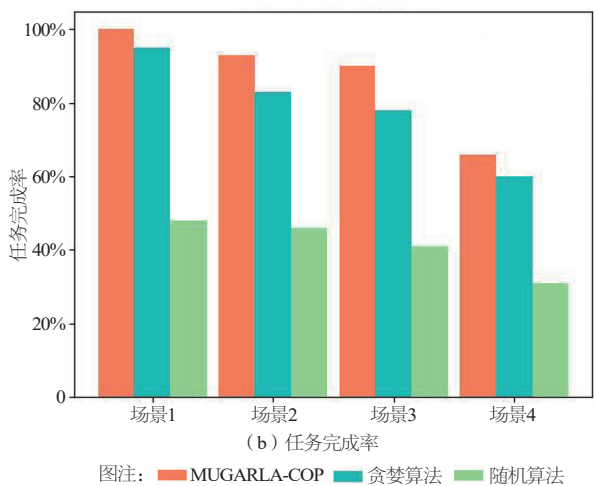
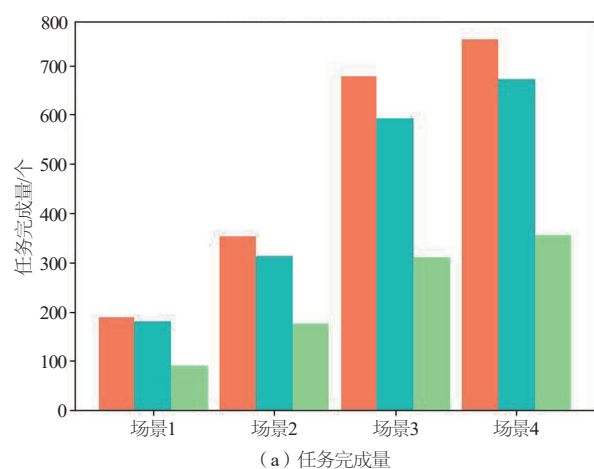


图7 场景1~场景4下的4种算法性能表现

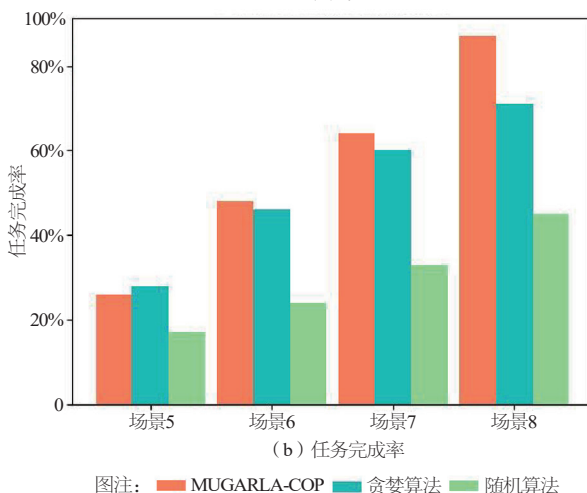
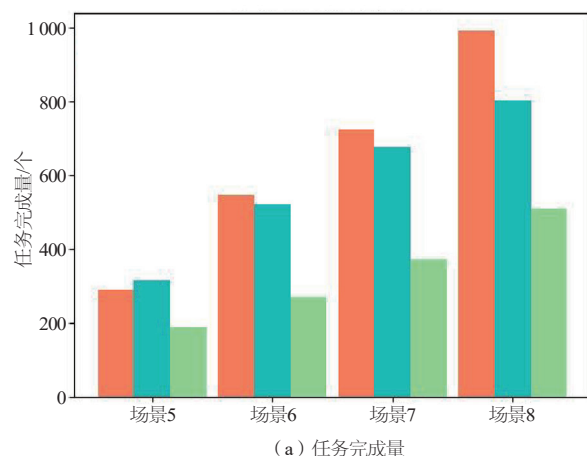


图8 4种地面站规模情况下3种算法的性能比较

率；随着规模的增加，测控任务间的冲突越来越大，尽管完成的任务数量不断增加，总体的任务完成率却在不断下降，MUGARLA-COP的任务完成情况较其他两者更有优势。

图8是4种地面站规模情况下3种算法的性能比较。场景5~场景8中各包含47个、83个、128个以及162个地面站。系统中包含1 134颗卫星。24 h内每颗卫星随机产生0~2个测控需求。

可见在地面站数量较少、测控资源相对有限的情况下，对于大规模星座测控任务的完成率较低。随着地面站个数的增加，系统可利用的测控资源增多，3种算法所能完成的任务需求数量也在提升，其中提升最显著的是MUGARLA-COP算法。这是因为随着地面站个数的增加，地面站Agent可行动作空间有限的劣势得到弥补，整个系统可以在更大的范围内遍历可行动作，从而能使算法收敛至较好的结果。

整体看来，在各种场景下，MUGARLA-COP算法的有效性优于其他两种算法。其中当地面站规模一定的时候以及当卫星数量和测控需求超过一定规模的时候，各种算法的任务完成数量只能得到很有限的提升，而任务完成率却不断下降。系统的卫星规模一定时，随着地面站个数的增加，能够完成的任务数量不断增加，任务完成率也在一定程度上改善。其中MUGARLA-COP算法的性能对地面站数量比较敏感，在地面站数量充足的情况下，能够依赖地面站Agent之间的交互实现较好的规划结果。

参考文献：

- [1] FOUST J. SpaceX's space-internet woes: despite technical glitches, the company plans to launch the first of nearly 12, 000 satellites in 2019[J]. IEEE Spectrum, 2019, 56(1): 50-51.
- [2] DEL PORTILLO I, CAMERON B G, CRAWLEY E F. A technical comparison of three low earth orbit satellite constellation systems to provide global broadband[J]. Acta Astronautica, 2019, 159: 123-135.
- [3] AKYILDIZ I F, KAK A. The Internet of space things/ CubeSats[J]. IEEE Network, 2019, 33(5): 212-218.
- [4] KAK A, AKYILDIZ I F. Designing large-scale constellations for the Internet of space things with CubeSats[J]. IEEE Internet of Things Journal, 2021, 8(3): 1749-1768.
- [5] SAAD W, BENNIS M, CHEN M. A vision of 6G wireless systems: applications, trends, technologies, and open research problems[J]. IEEE Network, 2020, 34(3): 134-142.
- [6] KATO N, MAO B, TANG F, et al. Ten challenges in advancing machine learning technologies toward 6G[J]. IEEE Wireless Communication, 2020, 27(3): 96-103.
- [7] KHAN L U, YAQOOB I, IMRAN M, et al. 6G wireless systems: a vision, architectural elements, and future directions[J]. IEEE Access, 2020, 8: 147029-147044.
- [8] KODHELI O, LAGUNAS E, MATURO N, al. Satellite communications in the new space era: a survey and future challenges[J]. IEEE Communications Surveys & Tutorials, 2021, 23(1): 70-109.
- [9] EVANS J V. Satellite systems for personal communications[J]. Proceedings of the IEEE, 1998, 86(7): 1325-1341.
- [10] 李军. 空天资源对地观测协同任务规划方法[D]. 长沙: 国防科学技术大学, 2013.
- LI J. Coordinated task planning research of space-aeronautics earth-observing[D]. Changsha: National University of Defense Technology, 2013.
- [11] 陈星. 统一扩频测控体制捕获技术的研究[D]. 长沙: 国防科学技术大学, 2010.
- CHEN X. Research on the acquisition technologies in spread spectrum TT&C system[D]. Changsha: National University of Defense Technology, 2010.
- [12] 易克初, 李怡, 孙晨华, 等. 卫星通信的近期发展与前景展望[J]. 通信学报, 2015, 36(6): 161-176.
- YI K C, LI Y, SUN C H, et al. Recent development and its prospect of satellite communications[J]. Journal on Communications, 2015, 36(6): 161-176.
- [13] CLEMENT B J, JOHNSTON M D. The deep space network scheduling problem[C]//Proceedings of the 17th conference on Innovative applications of artificial intelligence - Volume 3. New York: ACM, 2005: 1514-1520.
- [14] ARBABI M, GARATE J A, KOCHER D F. Interactive real time scheduling and control[C]//Summer Computer Simulation Conference. [S.l.:s.n.], 1985.
- [15] 金光, 武小悦, 高卫斌. 卫星地面站资源调度优化模型及启发式算法[J]. 系统工程与电子技术, 2004, 26(12): 1839-1841.
- JIN G, WU X Y, GAO W B. Ground station resource scheduling optimization model and its heuristic algorithm[J]. Systems Engineering and Electronics, 2004, 26(12): 1839-1841.
- [16] 张娜, 柯良军, 冯祖仁. 一种新的卫星测控资源调度模型及其求解算法[J]. 宇航学报, 2009, 30(5): 2140-2145.
- ZHANG N, KE L J, FENG Z R. A new model for satellite TT & C resource scheduling and its solution algorithm[J]. Journal of Astronautics, 2009, 30(5): 2140-2145.

- [17] 王远振, 赵坚, 聂成. 多卫星—地面站系统的Petri网模型研究[J]. 空军工程大学学报(自然科学版), 2003, 4(2): 7-11.

WANG Y Z, ZHAO J, NIE C. Study on Petri net model for multi-satellites-ground station system[J]. Journal of Air Force Engineering University (Natural Science Edition), 2003, 4(2): 7-11.

- [18] 刘洋, 陈英武, 谭跃进. 卫星地面站系统任务调度的动态规划方法[J]. 中国空间科学技术, 2005, 25(1): 44-47.

LIU Y, CHEN Y W, TAN Y J. The method of mission planning of the ground station of satellite based on dynamic programming[J]. Chinese Space Science and Technology, 2005, 25(1): 44-47.

- [19] 杜红梅, 柯宏发. 基于多智能体技术的航天测控资源调度模型设计[C]//系统仿真技术及其应用(第16卷). 2015: 78-81.

DU H M, KE H F. Design of space TT&C resource scheduling model based on multi-agent technology [C]//System Simulation Technology and Its Application. [S.l.:s.n.], 2015: 78-81.

- [20] 韩伟, 张学庆. 一种基于离散粒子群的多星任务规划算法[J]. 无线电工程, 2015, 45(1): 1-4.

HAN W, ZHANG X Q. A multisatellite task planning algorithm based on discrete particle swarm[J]. Radio Engineering, 2015, 45(1): 1-4.

- [21] 王海波, 徐敏强, 王日新, 等. 基于蚁群优化-模拟退火的天地测控资源联合调度[J]. 宇航学报, 2012, 33(11): 1636-1645.

WANG H B, XU M Q, WANG R X, et al. Solving space and ground TT & C resources integrated scheduling problem with ant colony optimization-simulated annealing algorithm[J]. Journal of Astronautics, 2012, 33(11): 1636-1645.

- [22] TAN M. Multi-agent reinforcement learning: independent versus cooperative agents[C]//Proceedings of the Tenth International Conference on International Conference on Machine Learning. New York: ACM, 1993: 330-337.

[作者简介]



刘阳 (2000-), 男, 西安电子科技大学本科生, 主要研究方向为空间信息网络任务规划方法。



周笛 (1991-), 女, 博士, 西安电子科技大学副教授, 主要研究方向为空间信息网络任务规划及资源管理、卫星互联网资源管控技术等。



盛敏 (1975-), 女, 博士, 西安电子科技大学教授, 主要研究方向为B5G/6G移动通信网络、异构网络融合以及空间信息网络等。



李建东 (1962-), 男, 博士, 西安电子科技大学教授, 主要研究方向为空间信息网络、智能无线网络、大规模自组织网等。



郝时光 (1986-), 男, 硕士, 现就职于中国空间技术研究院, 主要研究方向为通信卫星载荷设计、空间组网系统设计。



郑晓天 (1986-), 男, 博士, 现就职于中国空间技术研究院, 主要研究方向为通信卫星载荷设计, 空间组网系统设计。