**基于强化学习的联合抗干扰技术**

# 系统模型

无人机集群通过搭载日益丰富的载荷能够执行侦察、打击、通信等多种军用或民用任务，因此得到广泛研究。在军事场景的强对抗环境中，针对无人机通信能力的干扰方式也逐渐增多，于是提高无人机集群应对复杂干扰的能力变得更加重要。

**现有的针对无人机集群抗干扰研究较少，其不足主要表现为以下两个方面**：一、相关研究的抗干扰场景相对简单；二、场景采用的抗干扰手段单一。针对第一个方面，现有的研究场景主要存在两个问题：**首先**是当研究轨迹控制和功率控制作为抗干扰主要方式时，其设置的场景下只存在基站/传感器与单架无人机的单条通信链路， 未考虑无人机集群情况下多架无人机之间的通信链路关系以及联合轨迹控制等复杂问题，无法处理无人机集群下的轨迹规划问题。**其次**是当以信道选择作为抗干扰主要方式时，其设置的场景下无人机保持静止并且干扰源一般只设置为单一扫频干扰，无人机之间的信道选择决策被简化，也就导致无法应对目前强对抗干扰环境下的适应性强的智能干扰场景。针对第二个方面，当前的研究一般只解决单一场景下的单一抗干扰手段问题，而解决多无人机在智能干扰源场景下执行任务的问题需要将信道选择、轨迹与功率控制结合进行联合抗干扰，才能充分利用无人机集群高灵活性的特点提高无人机集群的抗干扰能力。

在一类无人机任务场景中，无人机需要执行**中继通信部署、目标区域数据收集**等任务，这需要无人机集群中的每架无人机需要**在各自的任务区域内保持活动**，并且每架无人机与无人机之间需要保持正常的通信，**实现实时的数据交互与指令传达等任务**。具体来说，针对中继通信，每架无人机需要部署在对应的通信受限地面的空中，用于实现接收地面用户的信息并中继转发给其他无人机，最终到达基站处；针对数据收集任务，每架无人机需要在侦察区域利用多种传感器持续采集对应信息，然后再通过无人机之间的链路回传给基站。

在这类场景中，每架无人机具有运动能力但无法离开自身的任务区域。设置场景中的无人机采用窄发宽收的方式进行通信，每架无人机选择自己的信道进行通信。与此同时智能干扰源出现在场景中的特定区域，能够自主决策移动以及改变自身的干扰方式，自适应地干扰无人集群之间的数据传输。无人机集群可以通过信道选择、轨迹和功率控制的方式进行联合抗干扰决策，从而提升场景中每架无人机之间的通信链路质量。

## 任务场景概述

设定问题所研究的场景为：若干无人机分别在场景的各个区域内执行任务，无人机节点需要在特定范围停留内以保证任务的执行，与此同时每架无人机之间存在通信交互的需求。每架无人机采用窄发宽收的机制，能够选择多个信道中的一个信道进行信号发送，并监听多个信道的接收信号。场景中存在一个或多个可以移动的智能干扰机，该干扰机的发射总功率固定并且能够将自身的功率自适应分配于多个信道上。**该场景需要解决的问题为**：无人机通过自主选择合适的发射信道、发射功率和移动位置，实现针对智能干扰源的联合抗干扰，以获得无人机与无人机之间最大的通信传输速率。



图 1‑1 无人机集群联合抗干扰场景示意图

## 场景数学建模

针对以上场景首先进行数学建模。在场景中，无人机的位置表示为,其中将仿真时间分为个时隙，表示对应的离散化时隙。

在该场景中，无人机处在自身初始节点周围一定范围内执行任务，节点约束为不超出该初始位置的一定距离内，对应的干扰源在其初始位置内的一定范围内移动，每次移动的距离在和方向上的位移为-1到1之间的可选值，对应的位置关系满足如下条件：

其中为节点的初始位置，是一个常数，表示为节点不能超过初始位置的距离。对场景中对每个节点来说，假设无人机节点之间的传播视为LOS视距传播，其对应的节点和节点之间的信道增益公式可以表示为：

其中代表在时的信道增益，代表在第个时隙的节点和节点之间的距离，和分别代表第个节点和第个节点的二维坐标（假设所有节点和干扰的高度一致），同样的有节点和干扰之间的信道增益公式为：

其中为干扰器的位置。因此根据香农定理，在节点和节点之间传输速率为：

其中表示噪声功率，代表第个节点在第个时隙的发射功率，[n]表示第个节点所在的信道，代表干扰在第个节点所在信道的干扰功率。对应的节点和干扰所在的信道以及对应的发射功率表示为如下：

其中代表无人机节点的最大发射功率，代表为场景中所有的子信道数目，为干扰器在所有信道发射功率的总和，表示节点n选择的信道序号。可以推出第个节点接收的传输速率总和为:

由此可以得出在第个时隙所有节点的传输速率总和表示为如下公式：

则对于无人机的优化问题建模表示为以下：

同样的对于干扰源的优化问题建模表示为如下：

# 基于逐步凸近似的优化求解分析

本节对上述优化问题进行进一步分析。

首先针对无人机抗干扰问题进行研究，我们假设干扰源信息（包括路径轨迹、功率）已知，单独求解在给定干扰源信息的情况下无人机的优化问题，同时由于式15对应的优化问题涉及无人机节点功率、轨迹和信道三个变量的优化，而我们在当前部分内容主要讨论轨迹优化问题求解，因此将上述建模问题进行如下简化：

其中代表第个节点的发射功率，代表干扰功率，为干扰源的位置信息，三者都作为已知信息假设为已知常数，可得

则对于每架无人机的轨迹优化问题建模表示为以下公式：

可以将公式19表示为以下公式：

其中：

接下来2.1节将对该优化问题进行推导。

## 凸近似的推导与转化

首先由于该优化问题需要同时优化,我们采用**交替优化**的方式，每次值优化单个的，而对于其他的都当做常量处理，便于后续求导。于是上述的优化问题可以推导为优化。

根据公式20,该优化问题转化为两部分差的优化问题，是非凸函数的原因是和存在分母平方项，通过对进行对一阶数泰勒展开，即可获得其对应的下界凸近似，其对应的表达是如下所示：

是 时的原函数值， 由梯度公式计算得出，分为两部分，展开得到完整的一阶泰勒展开表达式：

其中对应的表达式为：

于是能够得到其上界，对于可以通过对其内部进行近似来求得其上界为对应值,引入对该式子进行放缩：

二次项非凸，对二次项一阶泰勒展开进行放缩：

于是得到：

由此得出：

故将优化问题凸近似为：

## 仿真求解结果与分析

根据公式32，使用SCA和交替优化的方式，步骤如下：

1. 初始化所有无人机的位置。

2. 对于每个迭代r=1,2,...:

a. 对于每个无人机l=1,...,M:

i. 固定其他无人机的位置。

ii. 构建当前无人机l的轨迹优化问题，应用SCA方法对目标函数进行凸近似。

iii. 使用CVX求解该凸优化问题，得到更新后的。

b. 检查收敛条件（如目标函数变化小于阈值或达到最大迭代次数）。

3. 输出最终的轨迹和总速率。

通过如上步骤，利用CVX公式即可得到各个场景下对应的迭代解结果。

**迭代求解结果待补充**

# 基于DDPG的联合抗干扰算法

## 网络模型架构

DDPG（深度确定性策略梯度）是一种能够同时处理离散和连续动作空间的深度强化学习算法，其核心架构融合了策略优化与价值评估的双重机制。整个框架围绕四个关键模块展开：Actor网络作为策略函数，直接输出确定性动作（如机械臂的精确移动角度），将状态映射到具体行为；Critic网**络**则扮演价值评判者的角色，通过Q值量化当前状态-动作对的长期收益，为策略优化提供方向。为缓解训练不稳定性，DDPG引入了**目标网络**：Actor与Critic各自拥有结构相同的副本，通过“软更新”（缓慢同步参数）生成更稳定的目标Q值，避免因网络突变导致的震荡。与此同时，经验回放池持续存储智能体与环境交互的历史数据，训练时随机抽取小批量样本打破数据相关性，使学习过程更接近独立同分布假设。这种架构通过Actor-Critic的协同迭代，成为复杂动态环境下强化学习应用的经典范式。

**本研究采用DDPG架构实现无人机联合抗干扰的智能决策输出，其中重点在于设计Actor网络的决策输出方式。因此本研究重点介绍适应该无人机场景下的Actor网络设计，以提高网络训练的效果。**



图 3‑1 DDPG架构

## Actor网络模型

Actor网络接收输入状态向量，经过特征提取后需要同时输出信道选择，轨迹和功率控制三种不同维度的输出，因此是本网络架构设计的重点。该网络模型结合了多头输出机制和掩码添加机制，具有较强的多任务学习能力。**1.多头输出机制：**在该网络的Actor部分，每个节点通过多个头进行输出，分别对应不同的任务，如信道选择、功率选择和移动选择。**2.掩码机制：**为了防止多个节点选择相同的信道并发生冲突，网络使用了掩码添加机制。在Actor网络的前向传播过程中，每次选择信道时，会根据当前已选的信道为其他信道设置一个高负值掩码，确保它们不会被重新选择。这种掩码机制保证了信道选择的互斥性，避免了信道冲突，从而提高了网络的稳定性和可靠性。掩码会在每次决策后更新，确保下一轮选择时不会选择已经被占用的信道。通过这种方式，网络能够有效地解决多节点间的资源竞争问题。

其中对应的Actor网络架构如图所示：



图 3‑2 Actor网络架构

如图，**该网络结构基于多头神经网络（Multi-Head Network）设计，目的是并行处理三种不同的决策任务：运动控制（Motion）、信道选择（Channel）和功率控制（Power）。该架构通过独立的头部分别生成这些决策，每个头部包含 Gumbel-Softmax、BatchNorm 等技术来优化训练过程，确保训练过程的稳定性和高效性。主要有以下几点设计**：

**1. 多头输出不同的决策（运动、信道、功率）：**网络通过多个独立的头部（例如 Motion-head、Channel-head、Power-head）生成不同的决策任务：运动控制（Motion-output）：输出对应无人机节点离散空间的选择；信道输出（Channel-output）：负责生成信道选择的决策，为每个无人机分配对应信道；功率输出（Power-output）：负责生成功率分配的决策。 每个决策任务都由独立的头部处理，这些头部共享来自主干网络的状态信息，但每个头部执行自己的任务。每个头部的网络结构相似，但会根据任务的不同生成特定的输出。为了确保每个任务的决策过程能够得到充分优化，Gumbel-Softmax 和 BatchNorm 被应用到每个任务的网络中。

**2. 信道选择决策通过mask机制防止碰撞：**在信道输出部分，网络使用 mask 机制来防止在信道选择过程中发生冲突或“碰撞”。具体来说，mask 是一种用于屏蔽不可能选择的信道的机制，确保网络在决策时只考虑有效信道，避免无效选择影响最终决策。mask 是一个与信道输出概率分布相同形状的张量，其值为0或1，表示信道是否可选。其主要通过以下方式实现：

①生成 mask：在第一个channel-head 内部，基于输入的状态信息（如干扰位置、功率分配等）生成 mask。这一过程会根据实际情况（例如干扰、功率限制等）决定哪些信道是有效的，哪些信道是不可以选择的。②**跨头传递 mask**：由于网络使用多头结构，mask 需要在不同的head 之间传递，确保每个头部都能基于一致的信息进行决策。例如，mask 可以从一个头部（如 Head\_1）传递到下一个头部（如 Head\_2），确保信道选择决策的一致性。mask 会在各个头部之间进行组合，使得所有头部的输出都只关注之前头部决策后的有效信道，从而降低决策输出空间的维度，防止维度爆炸的出现。

**③**决策生成：在 mask 过滤后的信道输出上，使用 Gumbel-Softmax 或最大概率选择生成最终的决策。mask 的作用确保了只有有效信道被考虑进最终的决策中，从而避免自碰撞信道影响网络的决策结果。

3. **BatchNorm 防止 Gumbel-Softmax 溢出：**每个头部在输入到softmax层之前都使用了 BatchNorm 层，这一技术有助于对输入进行标准化，防止数值溢出和梯度爆炸。BatchNorm 的基本工作原理为：对每一层的输入进行标准化，计算当前 mini-batch 的均值和方差；使用均值和方差对输入进行调整，使得输出均值为零、方差为单位，从而避免激活值的极端变化；通过学习 gamma 和 beta 参数来进行缩放和平移操作，保证网络的稳定性。此时，BatchNorm 通过对输入进行标准化，避免了这种数值不稳定，防止梯度消失或爆炸，从而保证网络训练过程的稳定。

**4. Gumbel-Softmax 实现 Soft Output 梯度回传，Hard Output 实现决策输出：**

Gumbel-Softmax用于解决离散决策问题，它通过将离散输出的概率分布平滑化，使得梯度能够顺利回传，允许基于连续近似的离散决策进行优化。Gumbel-Softmax 的公式如下：

其中，是类别的概率分布，是 Gumbel 噪声，τ是温度参数。通过调整τ 值，Gumbel-Softmax 在软输出和硬输出之间进行平滑过渡。Hard Output：生成最终的离散决策输出，通过选择最大概率的类别来生成硬选择。硬选择会截断梯度，因此 Gumbel-Softmax 使用软选择的输出并通过温度τ 来平滑硬输出和软输出之间的过渡。温度控制（Tau）：tau 控制着 Gumbel-Softmax 的输出平滑度。较大的 tau 值使得输出更平滑，接近于连续分布，从而便于梯度回传。较小的 tau 值则使得输出更接近离散决策，减少软输出与硬输出之间的过渡。

该网络架构通过多头结构实现了轨迹、信道和功率三个不同决策任务的并行处理，每个任务通过独立的头部来生成输出。网络使用了 Gumbel-Softmax、BatchNorm 和 mask 等技术来优化训练过程，确保每个任务的决策过程稳定高效。mask 机制在信道选择中起到了重要作用，通过屏蔽之前头部的决策输出保证了网络只关注有效的决策空间。Gumbel-Softmax 则通过平滑的软选择和硬选择输出实现了离散决策问题的有效优化，并通过调整 tau 参数控制软决策与硬决策之间的过渡。BatchNorm 的应用防止了其溢出现象，确保了训练过程的稳定性。

## 模型训练与结果分析

通过设计GYM场景，仿真实现数据采集，其中输入的状态向量由无人机状态向量和干扰节点状态向量两者组成，其中无人机状态向量包括：所有无人机的当前时刻的位置、功率决策和信道占用状态；干扰机状态向量包括：输出的动作向量为无人机节点的联合决策：节点下一时刻的移动决策、功率控制和信道选择决策。通过经验回访池机制收集数据并进行训练，对应的数据格式如下所示：



图 3‑3 输入状态向量



图 3‑4 输出状态向量

根据第一章所设计的一类场景，以及第三章提出的网络架构，设计不同的场景参数进行网络训练与结果分析，其主要从场景中无人机节点数目、干扰机相对位置和干扰机功率分配方式三个不同方面，评估该算法的效果。

### 节点数目

通过设置不同节点规模，分析不同节点数目下该算法的表现与效果，分别设置节点数目为**4节点和6节点**两种不同节点场景，其余参数设置和第二节一致，节点保持网格均匀分布，干扰采用单扫频干扰，干扰节点匀速巡航。

首先是4节点场景，其训练500个epoch后测试效果入下，其中上图表示每个节点选择的信道和干扰节点干扰的信道：可以看出信道选择避开了干扰节点，在信道选择上实现了互相选择不冲突的信道并且能够和下一状态的预测的干扰机占用的信道避开，以获得通信质量最高；下图展示了无人机在测试时轨迹决策，由于是单扫频干扰，当无人机避开该干扰后，无人机通信不会受到干扰机干扰，无人机决策为保持距离相互最近，以获取最大通信质量。

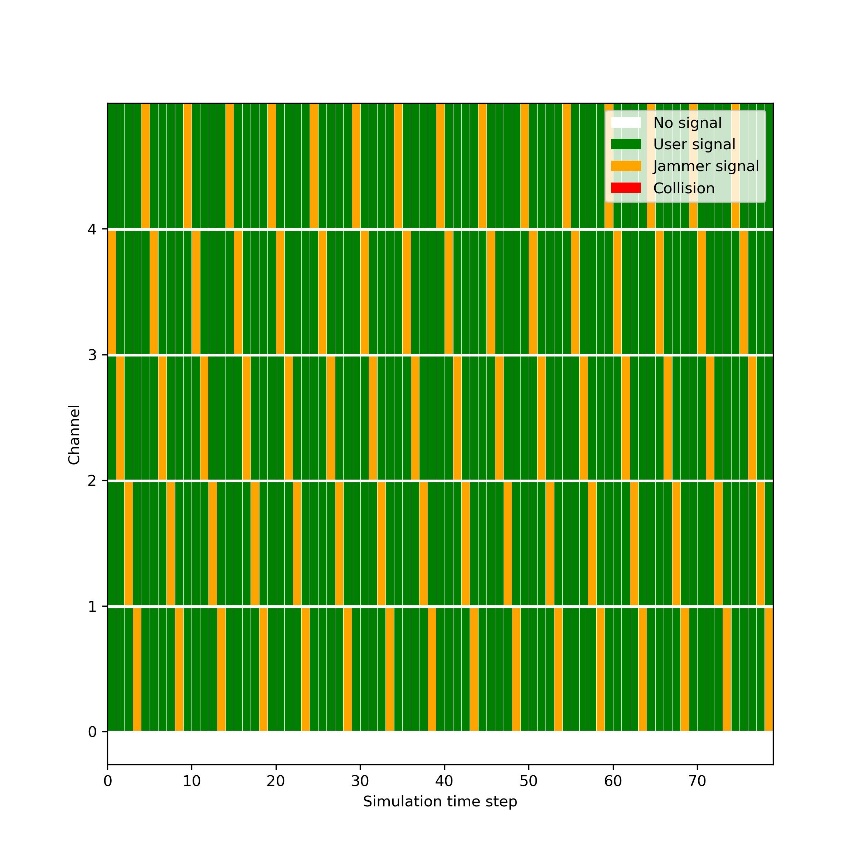


图 3‑5 4节点场景信道选择结果

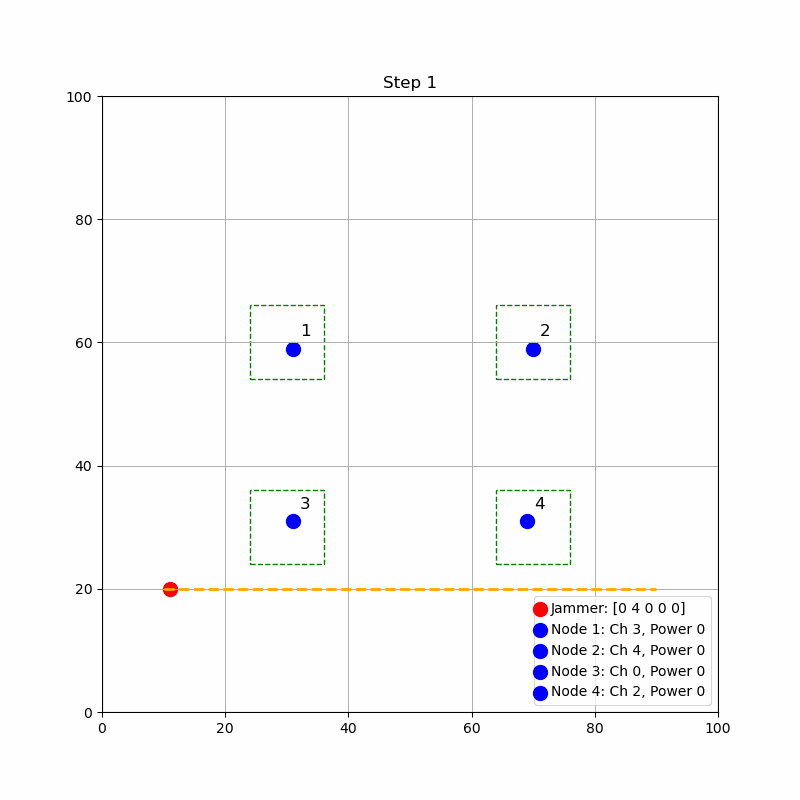


图 3‑6 4节点场景测试结果

然后是6节点场景，其训练500个epoch后测试效果入下，其中上图表示每个节点选择的信道和干扰节点干扰的信道：可以看出信道选择避开了干扰节点，在信道选择上实现了互相选择不冲突的信道并且能够和下一状态的预测的干扰机占用的信道避开，以获得通信质量最高；下图展示了无人机在测试时轨迹决策，由于是单扫频干扰，当无人机避开该干扰后，无人机通信不会受到干扰机干扰，无人机决策同样为保持距离相互最近，以获取最大通信质量。**但是实际效果中可以看出无人机决策并未收敛到最佳为止，网络模型还有调参优化的空间。**

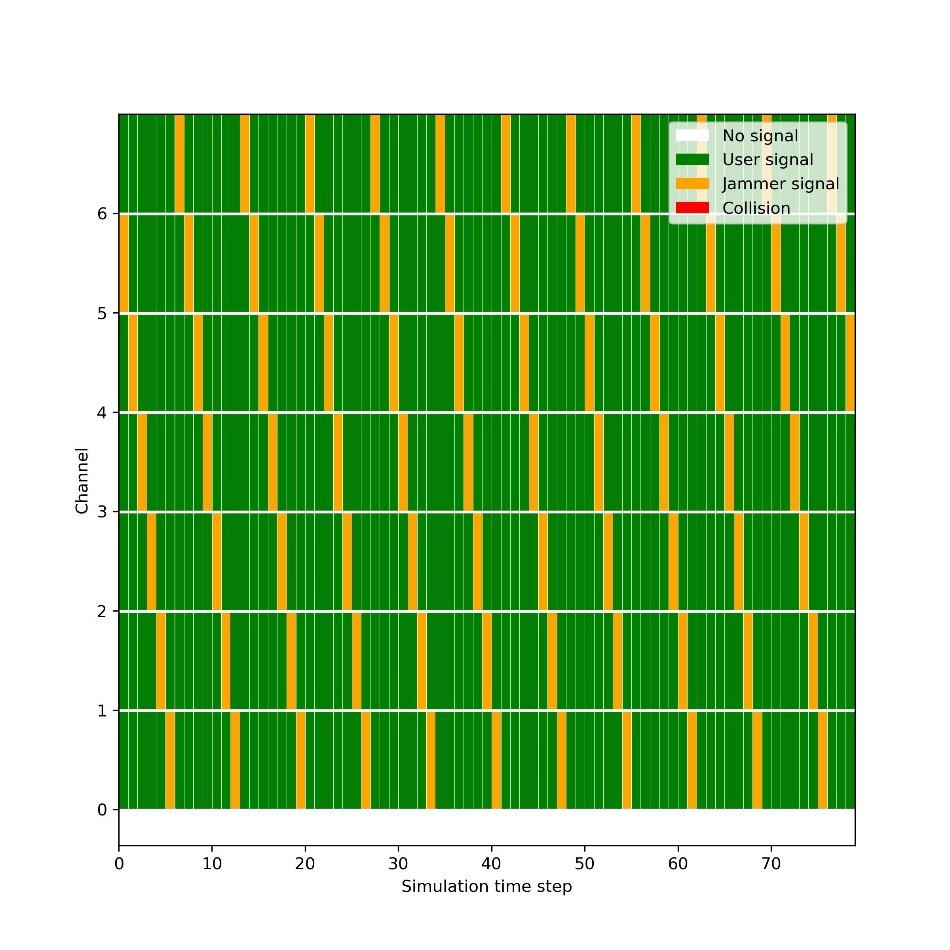


图 3‑7 6节点场景信道选择结果

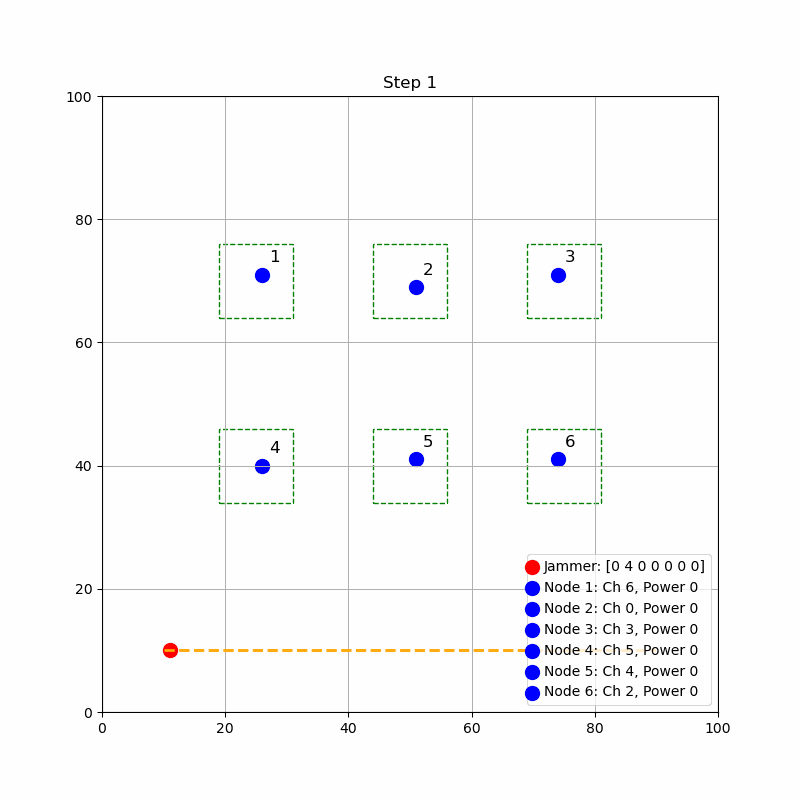


图 3‑8 6节点测试结果

### 按干扰模式分析

通过设置不同干扰模式，分析不同干扰功率分配模式下该算法的表现与效果，分别设置单扫频干扰、平均干扰和给定干扰模式，干扰处于节点中间区域，其余参数设置和第二节一致，节点保持网格均匀分布。

首先是单扫频干扰：在图可以看到节点总是选择未干扰的信道进行通信，由于节点不受到干扰机干扰，可以看出在图中节点选择相互靠近作为提高信道容量的最优解：

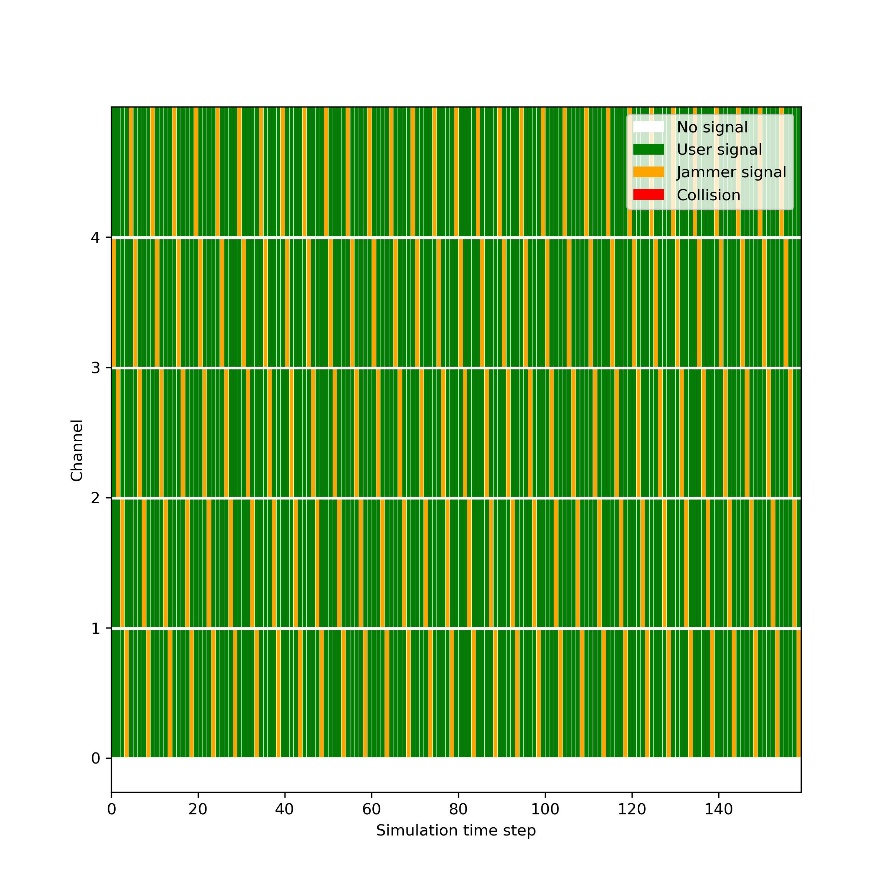


图 3‑9 单扫频干扰信道选择

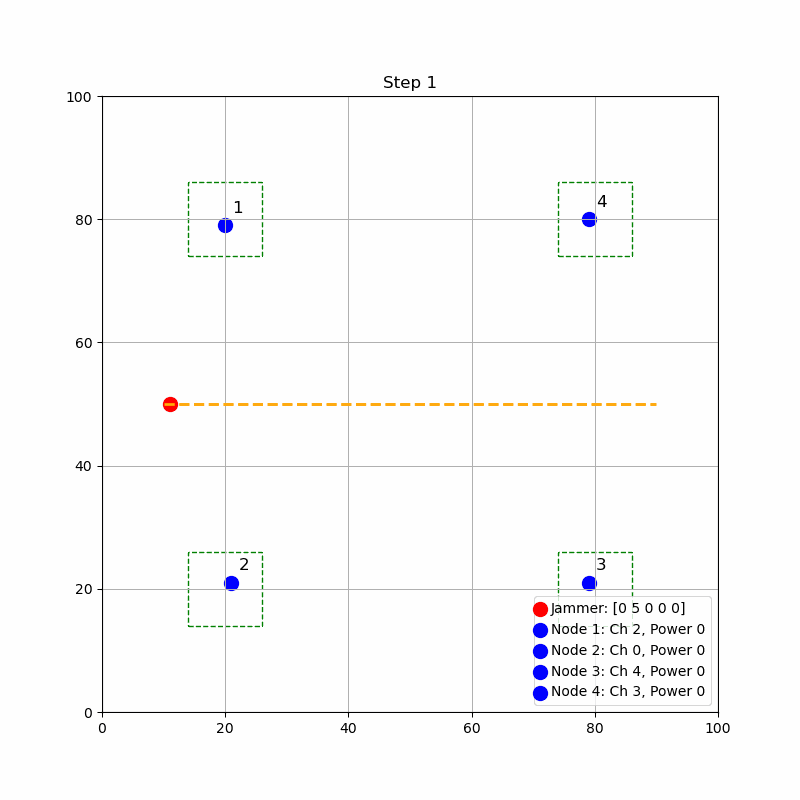


图 3‑10 单扫频干扰结果

然后是平均干扰，此干扰模式下，干扰机将自身的干扰功率平均分配给各个信道，此时无人机不管怎么选择信道，受到的干扰都相近，此时信道选择对抗干扰提升有限。无人机主要通过轨迹控制的方式，实现干扰机和无人机之间距离的最优关系，从而获得信道容量的提升，对应的结果如下：

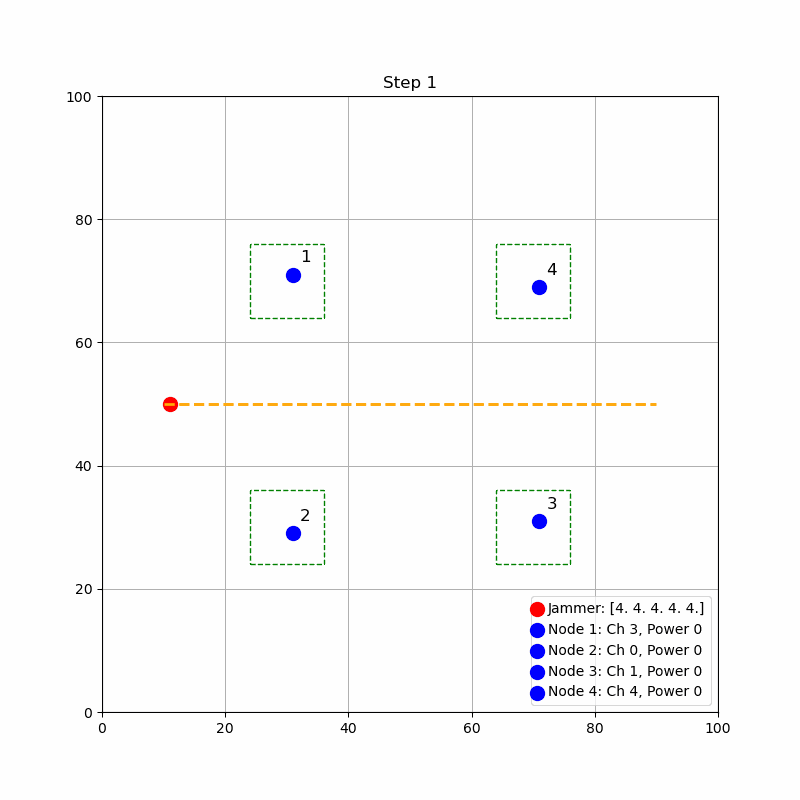


图 3‑11 平均干扰测试结果

最后是单扫频干扰加平均干扰，此干扰模式下干扰机在每个信道平均分配一定的干扰功率，此外还将多余的干扰功率以扫频的方式分配到某一信道上，既实现全信道干扰，也实现专门某一信道的重点干扰。网络训练结果如下，从图可以看出网络在信道选择方面能够成功避开功率更高的主扫频干扰，从而减少网络受到的干扰，在轨迹规划方面，节点能够随着干扰机位置的不同调整各自的位置，以实现在干扰情况下的信道容量提升。

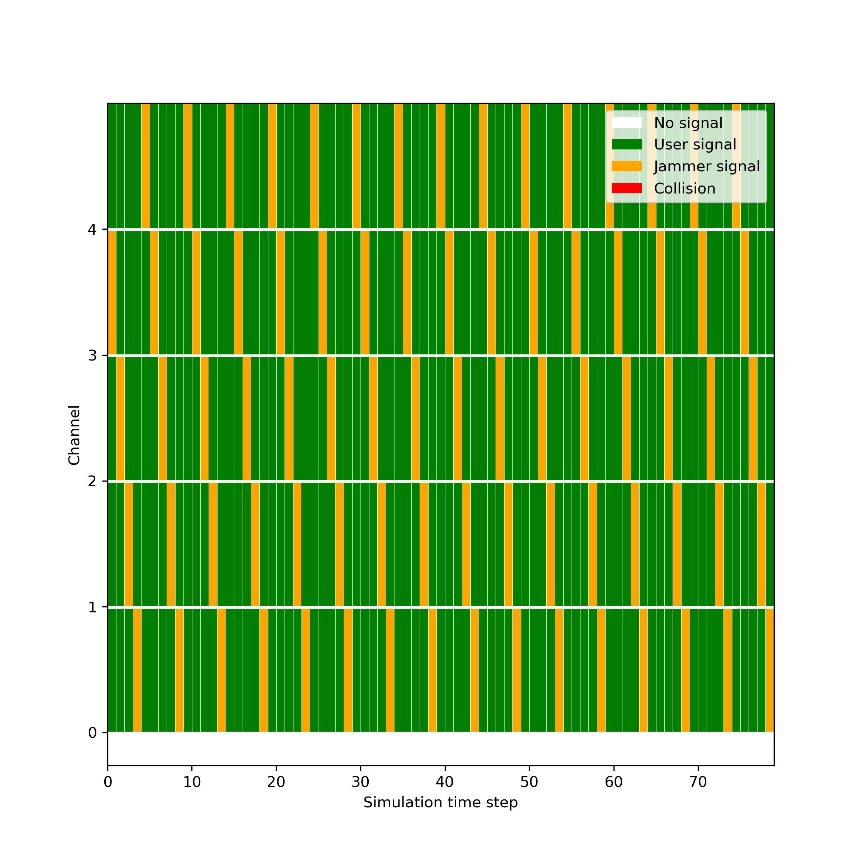


图 3‑12 单扫频干扰加平均干扰信道选择

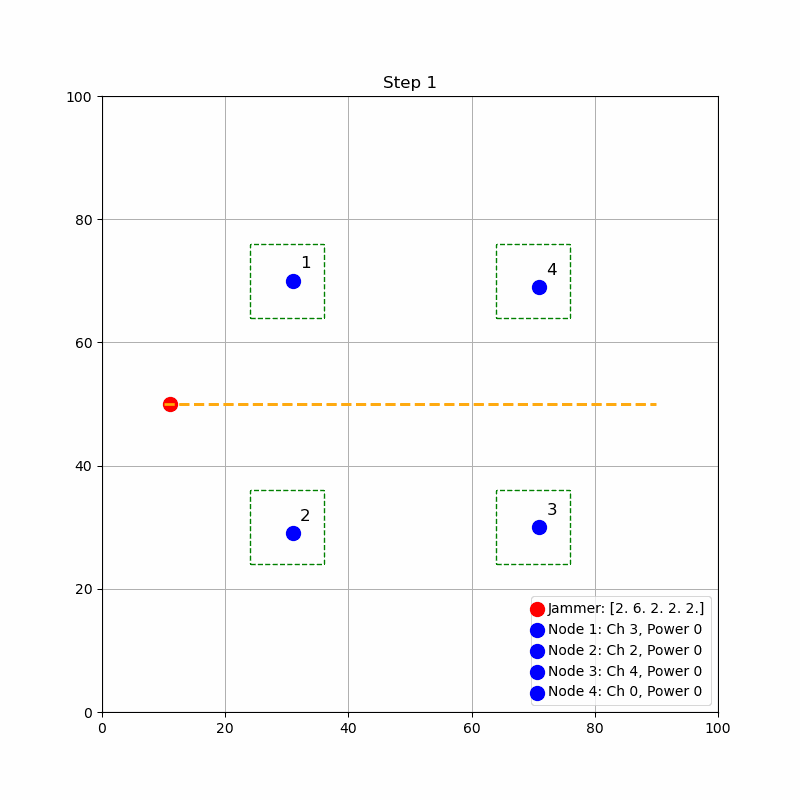


图 3‑13 单扫频干扰加平均干扰测试结果

### 按照干扰位置分析

通过设置不同的干扰所处的位置，分别将干扰设置在无人机任务区域中央和外部，研究不同干扰位置下该算法的性能。此场景下，干扰机采取平均干扰加扫频干扰的模式，节点保持网格均匀分布。  
 首先是干扰在外部，可以看出无人机节点总能避开主干扰信道，并且通过位移选择相互距离最近的位置以获取最大信道容量：

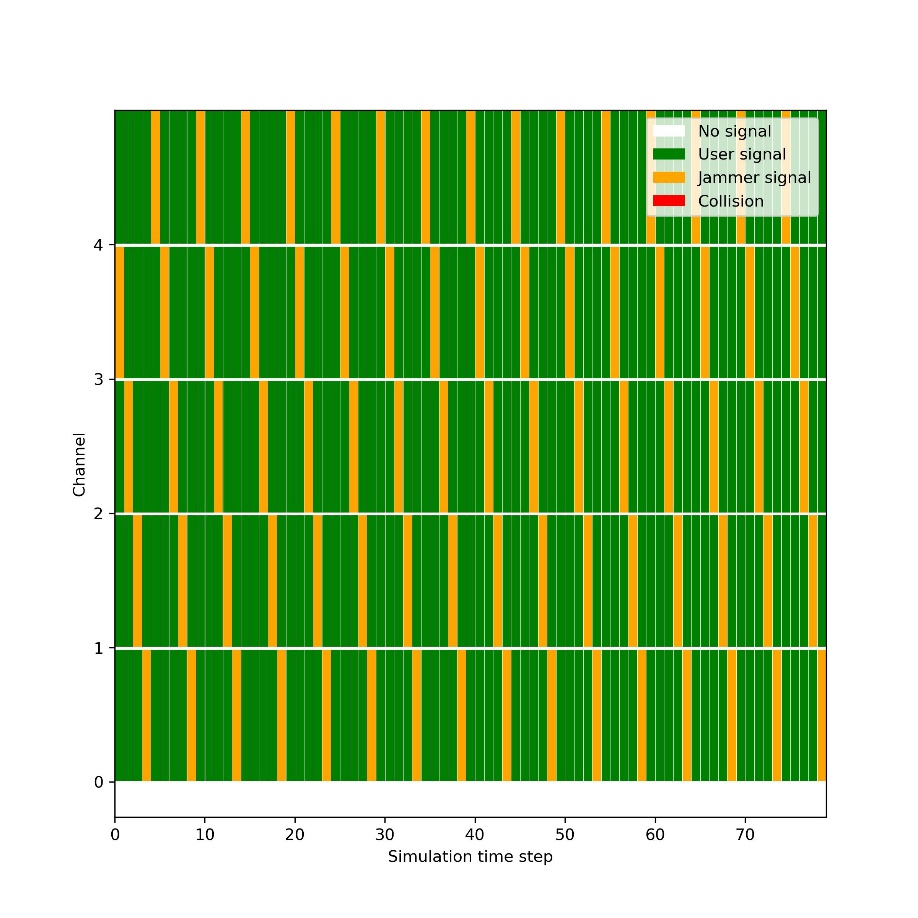


图 3‑14 单扫频干扰加平均干扰，干扰集群在外部信道选择

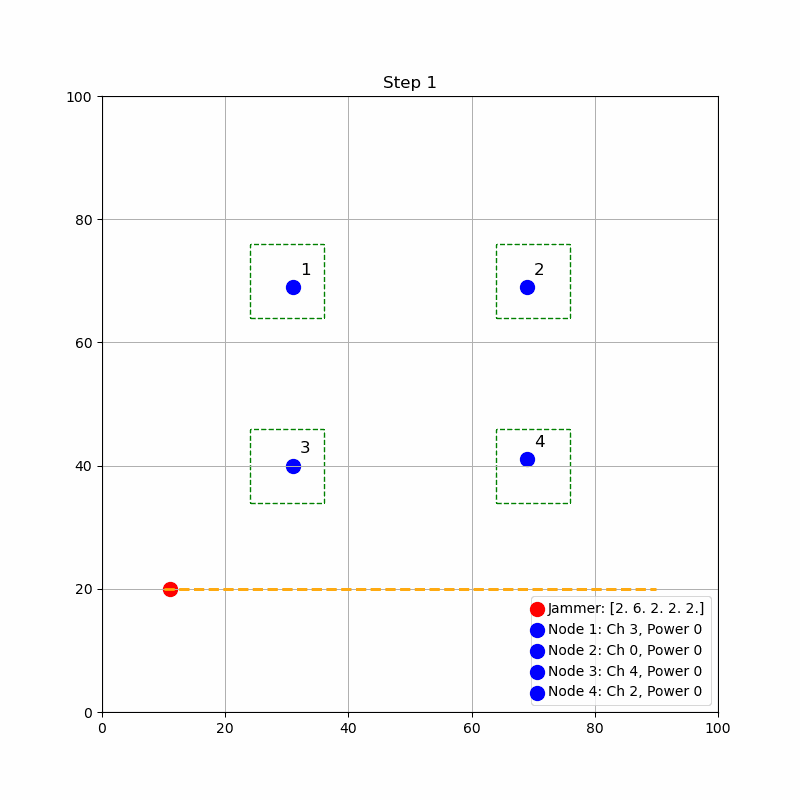


图 3‑15 单单扫频干扰加平均干扰，干扰在集群外部测试结果

其次是干扰在内部，该种情况下的效果在[3.2.2](#_按干扰模式分析)节有所讨论。

### 能量管理

添加能量管理模块作为限制功率的指标，实现联合功率决策输出，**待补充**。

### 智能干扰

设计智能干扰机，实现干扰机移动和干扰信道的智能化选择决策，实现无人机和干扰机的对抗，也是算法核心要解决的难点**。待补充。**

# 仿真结果对比分析

对比传统优化方法和本方法的不同，**待补充。**

# 研究创新点分析

1.设计了无人机集群任务场景，建模多智能体场景下的协同抗干扰，和之前研究较多的单一通信链路对场景下的路径规划场景相比，提出来了无人机集群路径规划场景下的建模与解决方案。

2.设计信道选择、功率和轨迹控制联合抗干扰决策，引入多智能体强化学习输出联合决策，能够提高应对复杂干扰场景能力。