# 基于强化学习的战术通信网频谱共享策略

赵旭 匡麟玲 倪祖耀 陆建华

（北京信息科学与技术国家研究中心）

摘要：随着现代战争对战术通信能力的需求不断提升，地面战术电台网与通信卫星互连互通逐渐成为战术通信发展的方向，而目前的战术电台网与卫星往往占用不同的频段以避免干扰，不存在频谱共享的模式。

当卫星的工作频度不高时，整个战术通信网的频谱利用率低，这驱动了战术电台网与卫星频谱共享的研究。本文着眼于这一思路构建了战术电台网与卫星的频谱共享模型，并基于马尔可夫决策过程构建决策模型，最终利用强化学习中的Q-learning方法通过训练得到其最优的频谱共享策略并就整体的频谱利用率以及电台网的通信稳定性进行了仿真比较。

关键字：战术电台网、通信卫星、频谱共享、马尔可夫决策过程、强化学习，Q-learning

**Abstract:** With the higher and higher demand of the tactical communication in the modern war, the terrestrial tactical radio network tends to connect with the satellite, in the way that they occupy separated spectrum bands without spectrum sharing to avoid the interference, leading to low spectrum efficiency when the satellite link does not work so frequently. Consequently, the spectrum sharing between the tactical radio network and the satellite is motivated. This paper formulate the spectrum sharing model among them and the corresponding decision model derived from Markov Decision Process model, after which the optimal policy is obtained by Q-learning method from Reinforcement Learning. Finally, simulation results are presented to verify the performance of the whole spectrum efficiency and the communication stability of the tactical radio network.

**Key Words:** tactical radio network, satellite communication, spectrum sharing, Markov Decision Process, Reinforcement Learning，Q-learning

# 介绍：

战术通信网作为现代战场通信系统的重要组成部分，包含了地域通信网、地面战术电台网、通信卫星等子系统，承担了战场信息保障的重要任务[[[1]](#endnote-1)]。随着当前战术通信的快速发展，地面战术电台网逐渐具备与卫星进行直接通信的能力，例如现在快速发展的各种战术通信卫星。相比于单纯的地面战术通信，“星地互连”大大地增强了战术通信的覆盖范围、系统容量、传输性能以及信息获取能力。

从业务特点上来说，地面战术电台网的战术通信（下文中简称为战术通信）相较于卫星通信有很大的不同。地面战术通信负责一个局部区域内各个节点之间的通信，受战场通信的需求影响其业务频繁，工作持续时间较长；卫星通信往往负责跨区域、跨军种、跨层级之间的通信，工作频度受业务特点决定，例如：当卫星传输总部的战役级指令或情报信息时，往往其工作频度比较低。而在实际设计当中，一般将战术通信与卫星通信划分到不同的频段以避免两者之间的相互干扰。因此，当卫星通信工作频度较低时，其频段会存在大量的空闲时隙，导致整个战术通信网的频谱利用率很低。

在此情况下，可以参考认知无线电中频谱共享[[[2]](#endnote-2)]的方法来提升频谱利用率，即将卫星视为主用户，战术电台网作为次用户能够在卫星不工作时利用其频谱，最终提高整体的频谱利用率。

另一方面，战术通信受地形遮挡、多径以及敌方干扰等不利因素影响[[[3]](#endnote-3)]，信道条件复杂多变，造成通信质量不稳定、易中断。对于较大带宽的信号传输，一般的物理层解决方法如跳扩频、编码、自适应速率调节往往都不能解决通信稳定性的问题[[[4]](#endnote-4)]。根据文章[[[5]](#endnote-5)]，在多信道传输的情况下，利用不同信道质量的差异性，通过合理地探测最佳接入信道，能够带来平均信道容量上的增益。因此，战术通信与卫星通信的频谱共享除了能够提升整体的频谱利用率之外，还可能提高战术通信的稳定性。

但是从实际上来说，卫星通信与战术通信的频谱共享存在着以下几个方面的问题：第一，卫星通信信号的易干扰性：由于卫星下行信号的强度在接收端非常微弱，因此如果战术通信与卫星通信的下行链路同时工作在同一频点，容易造成战术通信对卫星下行链路的干扰；第二，战术通信链路之间的同频干扰：由于战术电台网中会存在不同的节点同时工作的情况，因此不同的链路可能同时接入卫星的频谱进而可能造成彼此之间的同频干扰；第三，频谱切换的代价：每一次的频谱切换过程都要伴随收发双方握手、载波同步等过程，因此会带来时间开销，频繁的频谱切换会严重制约系统的通信性能。本文重点考虑前两点问题，暂时不考虑频谱切换开销的问题。

当前，关于军事通信中的频谱共享研究非常少，其主要原因在于安全性的考虑[[[6]](#endnote-6)]，但是本文所研究的战术电台网与卫星属于协作关系，同属于一个战术通信网，因此不需要考虑因彼此之间的频谱共享所带来的安全性的问题。除此之外，本文中频谱共享的目的不光在于提高频谱利用率，还在于提高战术通信的稳定性，这也不同于一般的认知无线电[[[7]](#endnote-7)]中频谱共享的目的。2016年在文章[[[8]](#endnote-8)]中首次研究了卫星通信与跳频数据链的频谱共享，第一次研究了战场通信中的频谱共享。该文从解决频谱资源紧张的目的出发，从物理层干扰的角度研究了卫星通信上行链路与数据链频谱共享的可能性，并且研究了数据链的最佳的跳频策略以减少与卫星通信的碰撞以及数据链相互之间的碰撞。本文则着重从接入层的角度研究卫星通信的下行链路与战术通信链路的频谱共享策略，并且专门针对战术通信中较大带宽信号的传输，不考虑跳频情况。

# 系统模型

## 频谱共享场景

本文假设存在一个战术通信系统，包括战术电台网以及通信卫星，分别对应战术通信与卫星通信。其中，战术电台网中的车载台具有双模工作能力，能够同时与战术电台与卫星进行通信。假设卫星通信的下行链路工作频带在 ，而战术电台网所工作的频带范围在 。对于战术电台网，假设同一时刻有个链路同时工作，每个链路占用一个提前划分的固定信道，每个信道带宽为，而卫星下行链路的带宽可以划分为个该信道，记为。频带的划分关系如图1所示：



图 1战术通信与卫星通信频谱划分

## 战术电台网通信模型

本文假设在研究的时间范围内，战术电台网中的工作链路集合不发生改变，并且每个链路均一直处于工作状态。由于在战术电台网中，各个节点均处于不断的移动当中，本文假设每个链路所对应的多普勒频率均为 ，并且每个链路的信道均经历Rayleigh衰落，根据文章[[[9]](#endnote-9)]中对Rayleigh信道的有限状态马尔可夫过程（FSMC）建模，本文假设不同的链路均经历独立特定的FSMC过程，假设接收信噪比的平均值为，状态数，则利用文章[5]的方法可以计算得到门限 及其转移概率 。若第个战术通信链路在第个时隙进行通信传输，接收端信噪比为，则若 ，其状态 。

在战术通信当中，本文假设存在中心控制节点以及控制信道。在每个时隙的通信过后，战术通信中的每个链路通过控制信道将自身的通信结果反馈给中心控制节点，中心控制节点可以通过控制信道将下一时隙的接入方案分配给各个链路。此外，本文假设中心控制节点还具备对卫星下行信号的频谱感知能力，即能够检测当前卫星下行链路的工作状态，因此，不妨假设中心控制节点为战术电台网中的车载台。

## 卫星通信模型

本文假设卫星通信的下行链路工作模式为“通断”两状态马尔可夫模型，其转移概率如下：

 （1）

即， 。

如果卫星的下行链路状态为，表明此时卫星的下行链路处于工作的状态，此时如果有战术通信的链路接入，则会对卫星产生干扰，如果此时状态为，则表明此时卫星的下行链路处于空闲状态，战术通信的链路可以接入。

严格来说，卫星的信道对于战术通信来说也是FSMC模型，其平均接收信噪比为，在本文中假设卫星的信道对于战术通信来说非常好，如 远大于 ，因此不妨假设若当前卫星链路处于空闲阶段，只要战术通信链路接入其信道且彼此无碰撞，则认为战术通信链路能够成功传输。

## 频谱共享框架

结合之前的模型假设，战术电台网与通信卫星的频谱共享框架如图2所示：



图 2战术通信与卫星通信频谱共享框架

其中，在第个时隙，控制节点根据当前卫星下行链路的感知结果 以及上一个时隙各个战术通信链路接入各自战术通信信道或者卫星通信信道对应的传输结果 来决定当前的接入策略 。其中，若，则第个战术通信链路接入其原来的战术通信信道，若，则第个战术通信链路随机选择卫星下行链路个信道当中的一个进行接入；若在战术通信信道成功传输则，否则，若在卫星通信信道成功传输则，否则。在这里，成功传输的标准设定为在一个时隙内接收信噪比大于解调信噪比门限。

# 最优频谱共享策略求解

考虑到实际的战术通信网的工作需求，本文以时隙吞吐量作为衡量指标，即平均一个时隙内战术通信网整体的成功传输数据量，包括战术通信以及卫星通信的传输。根据模型本身的马尔可夫动态变化特性，本文利用马尔可夫决策过程来求解对应的最优频谱共享策略。

## 基于马尔可夫决策过程的决策模型

马尔可夫决策过程（MDP）[[[10]](#endnote-10)][[[11]](#endnote-11)]通过构建状态、行为、决策、收益模型，利用系统状态转移以及回报函数的马尔可夫变化特性，得到最佳决策使得平均累积折损收益最大[7]。

具体的建模过程如下：

在第个时隙，设战术通信网的状态为：

 （2）

其中，为第个时隙感知到的卫星链路的状态； 为第个时隙个战术通信链路在各自信道通信传输的结果，其中； 为第个时隙个战术通信链路的属性状态，即若当前第个战术通信链路所接入的信道为其对应的战术通信信道，则；若为卫星通信信道，则。

行为定义为：

 （3）

其中，若，则第个战术通信链路接入自身分配的战术通信信道，若，则第个战术通信链路接入卫星通信信道。

收益定义为：

 （4）

其中，表示卫星通信链路的传输结果，即若卫星通信下行链路在工作时受到了战术通信的干扰则，若无干扰则，若卫星通信下行链路不工作，则 。

此外，可以利用以上信息求解状态转移概率 ，以及回报函数概率 ，受篇幅限制本文省略。

## 最优共享策略求解

本文中，假设战术电台网中控制节点不知道状态转移概率以及回报函数概率，在此情况下，根据强化学习[[[12]](#endnote-12)]，可以采用Q-learning的方法在模型未知的情况下求解最优的共享策略。其方法是通过如下迭代式完成。

 （5）

其中， 为更新因子，为折损系数。理论证明，当满足一定条件[11]且所有的状态行为组合都经历了无数遍时，Q函数值通过迭代能够收敛到最佳策略所对应的Q函数值，即满足， 有

 （6）

所对应的最佳策略为：对于

 （7）

# 仿真结果

在本文中，战术通信相关的仿真参数如下：链路数，多普勒频率，决策时隙，状态数[5]，时帧长度，信号带宽 ，调制方式为BPSK，

平均接收信噪比，解调信噪比门限 ；卫星通信的下行链路参数如下：信道数， ，（卫星业务到达间隔时间长，持续时间长）以及，（卫星业务到达时间间隔短，持续时间短） [[[13]](#endnote-13)]；Q-learning算法的参数如下：折损因子 ，更新因子 ， 为状态出现的次数，仿真长度80个时帧。

如果关注战术通信网整体的单位时隙吞吐量，其仿真结果如下图所示：

图 3 整体单位时隙吞吐量比较

可以看到不论卫星的业务特点是哪种，从战术通信网整体的单位时隙吞吐量来衡量，基于Q-learning的频谱共享策略要优于“STAY”策略以及“HOP”策略。其中，“STAY”策略指战术通信链路只利用本身的战术通信信道，对应于当前无共享的频谱使用模式，“HOP”策略指战术通信链路每次只占用卫星通信的信道，对应于完全共享的频谱使用模式。由于在基于Q-learning的频谱共享策略下，战术通信网能够利用相同的频谱达更大的单位时隙吞吐量，因此其对应的频谱利用率也更高。

另一方面，如果只关心战术电台网的通信稳定性，其中断概率在不同策略下的性能如图4所示。可以看到，在该频谱共享策略下，每帧的中断概率也比单纯地接入原战术通信信道要低，因此其通信稳定性也得到了提高。当然通过仿真也可以发现，由于不考虑对卫星干扰的影响，在该仿真条件下完全占用卫星通信的信道能够达到更低的中断概率。

值得强调的是，频谱共享策略之所以有效，其本质原因在于牺牲了卫星通信的性能，也就是以卫星通信受到了干扰为代价换取了战术通信系统整体频谱利用率的提升以及战术电台网通信稳定性的提高。



图 4 战术电台网中断概率比较

# 总结

本文针对地面战术电台网与通信卫星互连互通的情景下，整体频谱利用率低、战术电台网通信不稳定的问题，首次提出了战术通信链路与卫星通信下行链路的频谱共享模型，并建立了基于马尔可夫决策过程的决策模型。在此模型下，通过强化学习中的Q-learning算法得到最优的频谱共享策略。仿真结果表明，在不同的卫星下行链路工作模式下，所提出的频谱共享策略相比于当前无频谱共享的策略都能够达到更高的战术通信网整体频谱利用率以及地面战术通信的稳定性。

参考文献

1. 于全. 战术通信理论与技术[M]. 电子工业出版社, 2009. [↑](#endnote-ref-1)
2. Etkin R, Parekh A, Tse D. Spectrum sharing for unlicensed bands[C]// First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks. 2007. [↑](#endnote-ref-2)
3. 范喜全[1], 匡镜明[1]. 一种复杂环境下的战术通信信道仿真方法[J]. 系统仿真学报, 2008(9). [↑](#endnote-ref-3)
4. Optimal Frequency-Temporal Opportunity Exploitation for Multichannel Ad Hoc Networks[J]. IEEE Transactions on Parallel and Distributed Systems, 2012, 23(12):2289-2302. [↑](#endnote-ref-4)
5. Sabharwal A , Khoshnevis A , Knightly E . Opportunistic spectral usage: Bounds and a multi-band CSMA/CA protocol[J]. Networking IEEE/ACM Transactions on, 2007, 15(3):533-545. [↑](#endnote-ref-5)
6. Park J M , Reed J H , Beex A A , et al. Security and Enforcement in Spectrum Sharing[J]. Proceedings of the IEEE, 2014, 102(3):270-281. [↑](#endnote-ref-6)
7. Haykin S. Cognitive radio: brain-empowered wireless communications[J]. IEEE Journal on Selected Areas in Communications, 2005, 23(2):201-220. [↑](#endnote-ref-7)
8. Baek H , Lim J . Spectrum Sharing for Coexistence of Fixed Satellite Services and Frequency Hopping Tactical Data Link[J]. IEEE Journal on Selected Areas in Communications, 2016:1-1. [↑](#endnote-ref-8)
9. Zhang Q , Kassam S A . Finite-state Markov model for Rayleigh fading channels[J]. IEEE Transactions on Communications, 1999, 47(11):1688-1692. [↑](#endnote-ref-9)
10. Рыков, В. В. Dynamic programming and Markov processes[M]. [↑](#endnote-ref-10)
11. Thrun S, Littman M L. Reinforcement Learning: An Introduction[J]. IEEE Transactions on Neural Networks, 2005, 16(1):285-286. [↑](#endnote-ref-11)
12. Sutton R , Barto A . Reinforcement Learning:An Introduction[M]. MIT Press, 1998. [↑](#endnote-ref-12)
13. Zhao Q, Lang T, Swami A, et al. Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework[J]. IEEE J.select.areas Communication, 2007, 25(3):589-600. [↑](#endnote-ref-13)