

传输时限约束下的能量收集无线传感器网络多址接入优化

杨澳钦¹, 宫傲宇², 房婷¹, 邓磊³, 李强⁴, 张一晋¹

(1. 南京理工大学电子工程与光电技术学院, 江苏 南京 210094; 2. 洛桑联邦理工学院计算机与通信科学学院, 瑞士 洛桑 1015;
3. 深圳大学电子与信息工程学院, 广东 深圳 518060; 4. 鹏城实验室, 广东 深圳 518055)

摘要: 随着能量收集无线传感器网络在环境监测、工业自动化、战场侦察等实时通信场景的广泛应用, 多址接入需要同时考虑数据分组严格的传输时限特性以及传感器节点的能量收集特性。由于节点互干扰、传输紧迫性及能量储存度等因素的固有耦合, 此多址接入的设计和优化相比于仅需考虑数据分组到达特性的传统多址接入具有更大挑战性。首先, 设计各节点接入行为依赖于全局传输紧迫性和剩余能量的集中式接入协议; 然后, 考虑集中式接入具有难以承受的控制开销, 设计各节点接入概率仅依赖于本地传输紧迫性和剩余能量的分布式接入协议。以最大化网络吞吐率为优化目标, 使用马尔可夫决策过程对集中式接入分别进行考虑所有数据分组信息的完整建模和仅考虑队首数据分组信息的简化建模, 并基于逆向归纳算法求解两种建模的最优集中式策略; 最后, 使用分布式马尔可夫决策过程对分布式接入协议进行简化建模, 并基于马尔可夫策略搜索方法提出 ϵ -最优分布式策略。仿真结果验证了简化建模的有效性和所提出策略相较于其他策略的性能优越性。

关键词: 传输时限; 能量收集; 马尔可夫决策过程; 多址接入

中图分类号: TN911

文献标志码: A

doi: 10.11959/j.issn.2096-3750.2022.00283

Optimization of multiple access in the energy harvesting wireless sensor network with delivery deadline constraint

YANG Aoqin¹, GONG Aoyu², FANG Ting¹, DENG Lei³, LI Qiang⁴, ZHANG Yijin¹

1. School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
2. School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland
3. College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China
4. Peng Cheng Laboratory, Shenzhen 518055, China

Abstract: With the wide application of the energy harvesting wireless sensor network (WSN) in many real-time communication scenarios, such as environmental monitoring, industrial automation and battlefield surveillance, the multiple access of such WSN needs to take into account both the delivery deadline constraint of data packets and the energy harvesting dynamics of sensor nodes. Due to the inherent decoupling of interference, delivery urgency and remaining energy, the design and optimization of such multiple access are more challenging than that of traditional multiple access that only needs to take into account the packet traffic pattern. A centralized access scheme was designed with the access actions relying on the global knowledge of current delivery urgency and remaining energy. And then, to avoid the costly overhead in the centralized access, a decentralized access scheme was designed with the access probabilities merely relying on the local knowledge of delivery urgency and remaining energy. Under the objective of maximizing the network throughput, the centralized access schemes were formulated with complete and simplified knowledge as two Markov decision

收稿日期: 2021-11-22; 修回日期: 2022-06-24

通信作者: 张一晋, yijin.zhang@gmail.com

基金项目: 国家自然科学基金资助项目 (No.62071236, No.61902256); 中央高校基本科研业务费专项资金资助项目 (No.30920021127); 鹏城实验室重大攻关项目 (No. PCL2021A15)

Foundation Items: The National Natural Science Foundation of China (No.62071236, No.61902256), The Fundamental Research Funds for the Central Universities (No.30920021127), The Major Key Project of Peng Cheng Laboratory (No. PCL2021A15)

processes (MDPs), respectively, and the backward induction algorithm was used to obtain optimal centralized policies for these MDPs. Furthermore, the decentralized access was formulated with simplified knowledge as a decentralized MDP, and the Markov policy search was used to propose an ε -optimal decentralized policy. Simulations under a wide range of network configurations were provided to verify the effectiveness of the simplified modeling and demonstrate the performance advantage of the proposed policies.

Key words: delivery deadline, energy harvesting, Markov decision process, multiple access

0 引言

无线传感器网络 (WSN, wireless sensor network) 由大量部署在监测区域的传感器节点组成^[1], 网络中的传感器节点通常需要共享无线信道, 并遵循事先制定的多址接入协议向同一接入点上行传输所观测到的数据信息。随着 WSN 在环境监测^[2]、工业自动化^[3]、侦察^[4]等实时通信场景的广泛应用, 传感器节点往往需要在规定时限内向接入点上行传输观测信息, 而传感器节点通常使用电池供电并且从自然界收集能量的特性又将不可避免地对传输及时性造成负面影响。因此, WSN 的多址接入需要同时考虑严格传输时限特性^[5]以及能量收集特性^[6-7], 并由于节点互干扰、传输紧迫性以及能量储存度等因素的固有耦合, 其设计和优化相较于仅考虑数据分组到达特性的传统多址接入具有更大挑战性。

针对严格传输时限约束下的多址接入策略已有一系列工作。文献[8]以最大化网络可靠性为目标, 推导了应用于认知无线网络次用户广播的时隙 ALOHA 协议在时限约束下的最优静态接入概率。文献[9]在文献[8]的基础上推导了编码型时隙 ALOHA 在时限约束下的各类网络性能指标的闭合表达式, 并提出了使网络可靠性最大化的次优静态接入概率以及相应的编码参数。文献[10]进一步拓展了文献[8]的工作, 证明了接入点具有多包接收能力情况下时隙 ALOHA 协议在时限约束下最优静态接入概率的唯一性, 并提出了计算此接入概率的定点迭代方法。文献[11]推导了基于重传的时隙 ALOHA 在时限约束下的网络可靠性表达式, 并提出了在满足给定吞吐量要求的同时最大化网络可靠性的最佳重传次数。文献[12]推导了传输时限和帧同步数据流下的 ALOHA 网络最大渐进吞吐率。文献[13]结合离散时间吸收马尔可夫链和随机几何理论对传输时限下的大规模 ALOHA 网络性能进行了建模。尽管如此, 以上工作均未将队列等待时间纳入传输时限, 并采用相同且固定的接入概率, 因

此未考虑如何使得各节点根据传输紧迫性程度和节点竞争程度采用不同的接入概率以进一步改善网络可靠性。鉴于此, 文献[14]允许各节点根据当前时隙未收到的反馈信号对下一时隙接入概率进行加倍或减半, 但假设各节点队列中数据分组统一地以一固定周期到达, 且数据分组的传输时限均小于上述周期长度, 并且其接入概率的动态调整仅基于简单启发而远未达到最优。

另一方面, 针对考虑能量收集特性的多址接入亦已有广泛研究。文献[15]考虑各节点依赖一个具有恒定电源的接入点进行能量收集并同时需要向该接入点传输数据, 使用凸优化技术提出了最大化单个时隙平均吞吐率的最优集中式传输方案, 但假设各节点不能存储能量为将来所用, 因此仅基于单个时隙进行优化。文献[16]转而考虑上述模型中各节点能够存储能量为将来可用的情况, 基于马尔可夫决策过程 (MDP, Markov decision processes) 理论提出了使得长期平均吞吐率最大化的最优集中式传输方案。文献[17]基于博弈论方法提出了各节点采用相同策略假设下的最优分布式策略, 允许各节点根据自身状态决定接入概率。进一步, 针对各节点可采用不同策略的更广泛情况, 文献[18]基于分布式马尔可夫决策过程^[19] (Dec-MDP, decentralized Markov decision processes) 理论提出了使得长期平均吞吐率最大化的最优和近优分布式策略。尽管如此, 以上工作均理想化地假设各节点始终有数据需要发送, 而忽略了对数据流详细特征的考虑。

此外, 已有学者针对兼顾传输时限和能量收集特性的接入协议开展研究。文献[20]将队列等待时间纳入传输时限, 提出了一种近似分析网络可靠性的马尔可夫过程理论模型, 但仍然假设各节点在能量足够时采用相同且固定的接入概率。文献[21]针对认知无线网络次用户依托主用户信号进行射频充电的场景推导了次用户各类性能指标, 但假设数据分组仅在队列为空时以一定概率到达且仅考虑单个次用户采用固定接入概率的情况。文献[22]

基于移动节点从多个固定无线能量源进行能量搜集的场景,以节点位置、剩余能量以及剩余传输时间为状态,使用约束 MDP 求解了吞吐量约束下最大化可靠性的最优接入策略,但假设数据分组仅在队列为空时以一定概率到达且仅考虑单个移动节点的情况。

综上所述,目前尚无允许各节点根据当前传输紧迫性和剩余能量动态调整接入概率的多址接入设计和优化研究。针对此缺陷,本文对文献[8,10-11,18,20]的工作进行拓展,主要贡献在于:

①设计根据所有节点传输紧迫性和剩余能量制定各节点接入概率的集中式接入协议,以及仅根据自身传输紧迫性和剩余能量制定各节点接入概率的分布式接入协议;②使用 MDP 理论对集中式接入分别进行考虑所有数据分组信息的完整建模和仅考虑队首数据分组信息的简化建模,并基于逆向归纳算法针对这两种模型提出最大化网络吞吐率的最优集中式策略;③进一步使用 Dec-MDP 理论对分布式接入进行简化建模,并基于马尔可夫策略搜索方法提出最大化网络吞吐率的 ϵ -最优分布式策略;④仿真验证简化建模的有效性和所提策略相较于对照策略的吞吐率性能优势。

本文工作的创新和特色在于:①建模和优化首次全面考虑了节点互干扰、数据流到达过程、数据分组传输时限、数据队列长度、能量储存容量和能量收集过程等多种网络要素,并能够拓展至其他更复杂的同构或异构情况;②考虑队首数据分组信息对决策制定的主导作用,仅基于各节点队首数据分组(而不是所有数据分组)交付时间的马尔可夫特性推导各节点传输紧迫性的变化,从而降低了协议设计、建模和优化的复杂度;③基于 Dec-MDP 理论所获得的 ϵ -最优分布式策略不再局限于随机接入^[8-14,17,20]或确定性接入^[15-16]等特殊形式分布式策略,而是涵盖所有可能形式的分布式策略,从而增加了策略最优性的适用范围。

需要指出,集中式策略尽管由于开销巨大具有应用局限性,但其建模方法是分布式策略建模方法的基础,其性能亦可以作为分布式策略性能的上界,用于评估分布式策略设计的好坏。另外,本文尽管考虑最大化常用的性能指标网络吞吐率^[10,12,14-18,21](等价于最大化给定数据分组到达率下的网络可靠性^[11,20-22]),但本文工作亦可以支持其他优化目标。

1 系统模型

系统模型如图1所示,本节将介绍传输时限模型、能量收集模型,以及这些模型所构成通信场景下的集中式和分布式接入协议,最后介绍本文的优化目标。

1.1 网络模型

与文献[8-12,14-18,21]一致,考虑由一个接入点和标号为 $\{1,2,\dots,N\}$ 的 N 个节点组成的单跳 WSN。假设时间轴被划分为等长的标号为 $t \in \{1,2,\dots,T\}$ 的时隙,且接入点和各节点均能够准确定位到时间隙的边界,其中 T 为考虑的时隙总数。每个节点 i 在时隙 t 以接入概率 $0 \leq a_t^i \leq 1$ 向接入点发送占据一个时隙的数据分组。鉴于数据分组可由信道编码进行保护而不受信道错误影响,本文假设当且仅当一个数据分组传输不受其他数据传输干扰构成冲突时,此数据分组可被接入点成功接收。与文献[10]一致,鉴于数据分组传输紧迫性以及节能考虑,不考虑应答与重传机制。此通信场景常见于学校、医院的小型一跳 WSN,但也可基于常用的 WSN 簇拓扑^[23]应用于大型多跳 WSN 的簇内通信。

1.2 传输时限模型

假设在每个时隙 t 开始时刻,节点间数据分组到达过程相互独立,每个节点均以概率 λ 产生一个新的数据分组到达数据队列。定义到达每个节点的数据分组均具有严格的传输时限 D ,即数据分组在队列中被允许逗留的最长时间为 D 个时隙。受传输时限 D 约束,可见各节点的数据队列最多具有 D 个数据分组。进一步,定义一个数据分组在某一时刻的剩余允许逗留时隙个数为该数据分组的交付时间。因此,对于节点 i ,可以定义一组0、1字符串 $L_t^i \triangleq f_1^i f_2^i \dots f_D^i$ 表征队列所有数据分组的交付时间,规定只有存在一个数据分组交付时间为 l 时 $f_l^i = 1$ 。另外,定义 $d_t^i \in \{0,1,\dots,D\}$ 为时隙 t 开始时刻队列 i 队首数据分组的交付时间,这里 $d_t^i = 0$ 表示节点 i 的数据队列为空。

1.3 能量收集模型

显而易见,各节点从自然界收集的能量与能量源类型、环境因素以及节点能量收集器有关,在不同时刻应该具有不同的收集能量值。基于此特点,大部分场景(如振动能量源)下各节点的能量到达过程可以根据具体情况近似地简化为一个随机独

或队首数据分组交付时间 d_t^i) 及剩余能量 e_t^i 告知接入点, 并可以从接入点在下行控制信道的广播获知其他节点信息, 还能知晓接入点为各节点制定的接入策略。另外, 假设各节点在同步时隙仍然可以在数据传输信道传输数据分组和从自然界收集能量。显而易见, 同步时隙越多, 策略设计可以依赖的全局信息则越多, 策略性能则会越优异, 但同时也会增加控制信道开销, 导致更多的系统资源消耗。鉴于此, 本文将在理论分析部分首先考虑每个时隙均为同步时隙而具有有限实用性的集中式接入情况, 然后考虑仅第一个时隙为同步时隙而具有较好实用性的分布式接入情况, 并且在仿真部分比较这两种情况的仿真性能表现。

1) 考虑所有数据分组交付时间的集中式接入协议: 假设每个时隙 t 均为同步时隙, 即接入点在每个时隙 t 开始时刻准确知晓各节点所有数据分组交付时间及剩余能量, 并根据此信息执行由集中式接入策略 π_{cen^*} 制定的时隙 t 决策规则, 确定性告知每个节点是否可以发送数据分组, 即在每个时隙 t 开始时刻为每个节点选择接入概率 $a_t^i \in \{0, 1\}$ 。总而言之, π_{cen^*} 会基于全局信息确定性地让传输紧迫度更高而能量充足的节点优先发送。

2) 仅考虑队首数据分组交付时间的集中式接入协议: 与上述协议区别仅在于仅根据各节点队首数据分组交付时间及剩余能量执行由集中式接入策略 π_{cen} 制定的时隙 t 决策规则。

3) 仅考虑队首数据分组交付时间的分布式接入协议: 仅时隙 1 为同步时隙, 即接入点仅在时隙 1 开始时刻准确知晓所有节点队首数据交付时间及剩余能量, 并根据这些信息制定分布式接入策略 π_{dis} , 然后在时隙 1 将该策略和所知信息一并告知每个节点。在后续每个时隙 t , 各节点 i 只能根据当前自身队首数据分组交付时间以及剩余能量执行 π_{dis} 制定的时隙 t 决策规则, 即在后续每个时隙 t 开始时刻确定性地选择接入概率 $a_t^i, 0 \leq a_t^i \leq 1$ 。总而言之, 一个节点的自身紧迫性和剩余能量改变时, 其制定的自身发送概率也会改变, 会在评估自己比其他节点具有更高紧迫性且具有充足能量时使用更高的发送概率。其“分布式特性”体现为: 接入点仅在时隙 1 进行集中决策从而初始化各节点的传输行为, 而各节点在后续每个时隙是互相独立地进行接入。

1.5 优化目标

显而易见, 尽管 π_{cen^*} 的通信开销及所需信息最多而实用性有限, 但其最优性能可以作为其他策略 π_{cen} 、 π_{dis} 最优性能的上界, 用于评估 π_{cen} 、 π_{dis} 设计的好坏。

定义网络吞吐率为所有 T 个时隙内每个时隙成功传输的数据分组数量的期望, 即

$$G \triangleq E \left[\frac{1}{T} \sum_{t=1}^T \sum_{i=1}^N a_t^i \prod_{j \neq i} (1 - a_t^j) \right] \quad (2)$$

本文将以最大化网络吞吐率为优化目标, 对第 1.4 节所述不同情况下的 3 种策略 π_{cen^*} 、 π_{cen} 及 π_{dis} 进行优化。

需要指出, 如果各节点均不会因为能量耗尽而无法传输, 那么每个节点的剩余能量区别在所设计协议的建模和优化中将不会起作用。原因在于: 如果每个节点始终具有足够能量用于传输, 则其每个时隙的接入概率选择范围 (即动作空间) 将始终不受能量影响, 因此鉴于吞吐率本身与能量无关, 所设计策略仅会在每个时隙根据传输紧迫度为每个节点做出最优接入概率选择。

2 集中式接入策略优化

根据第 1 节所述系统模型, 本节将对严格传输时限下能量收集 WSN 的两种集中式接入协议分别进行有限时域 MDP 建模, 并基于此建模求解最优集中式策略。

2.1 MDP 完整建模

本节对第 1.4 节所述考虑所有数据分组交付时间的集中式接入协议进行 MDP 完整建模, 包括对其状态、行为、状态转移概率、决策规则、奖励以及期望总奖励的定义和推导。

状态: 考虑所有数据分组的交付时间, 定义时隙 t 的完整网络状态 $s_{\text{cen}^*, t}$ 为

$$s_{\text{cen}^*, t} \triangleq (L_t^1, e_t^1, L_t^2, e_t^2, \dots, L_t^N, e_t^N) \quad (3)$$

$s_{\text{cen}^*, t}^i \triangleq (L_t^i, e_t^i)$ 可视为时隙 t 时节点 i 的完整个体状态。通过列举所有可能的完整网络状态和完整个体状态, 可以分别得到完整网络状态空间 S_{cen^*} 和各节点 i 的完整个体状态空间 $S_{\text{cen}^*}^i$ 。

行为: 基于每个节点 i 在时隙 t 开始时刻的个体行为即接入概率 a_t^i , 定义时隙 t 的网络行为 a_t 为

$$a_t \triangleq (a_t^1, a_t^2, \dots, a_t^N) \quad (4)$$

由于能量充足且数据队列非空的节点可发送数据分组, 并且集中式接入可避免冲突, 因此 $\sum_{i=1}^N a_i^i \leq 1$, 并且当 $L_i \neq 00 \cdots 0$ 且 $e_i^i \geq m$ 时, $a_i^i \in \{0, 1\}$, 否则 $a_i^i = 0$ 。按此规则, 可以得到完整网络状态值为 $s_{\text{cen}^*, t}$ 时的网络行为空间 $A_{\text{cen}^*, s}$, 以及各节点 i 完整本地状态值为 $s_{\text{cen}^*, t}^i$ 时的节点 i 个体行为空间 $A_{\text{cen}^*, s}^i$ 。

完整状态转移概率: 用 $p_{\text{cen}^*, t}(s_{\text{cen}^*, t}^i, s_{\text{cen}^*, t}, a)$ 表示给定完整网络状态 $s_{\text{cen}^*, t} = s_{\text{cen}^*} \in S_{\text{cen}^*}$ 及网络行为 $a_t = a \in A_{\text{cen}^*, s}$ 时 $s_{\text{cen}^*, t+1} = s_{\text{cen}^*}$ 的条件概率。用 $p_{\text{cen}^*, t}^i(s_{\text{cen}^*, t}^i, s_{\text{cen}^*, t}^i, a^i)$ 表示给定节点 i 完整个体状态 $s_{\text{cen}^*, t}^i = s_{\text{cen}^*}^i \in S_{\text{cen}^*}^i$ 及行为 $a_t^i = a^i \in A_{\text{cen}^*, s}^i$ 时 $s_{\text{cen}^*, t+1}^i = s_{\text{cen}^*}^i \in S_{\text{cen}^*}^i$ 的条件概率。由于不采取重传机制, 可知各节点 i 完整个体状态的转移相互独立, 因此可以将 $p_{\text{cen}^*, t}(s_{\text{cen}^*, t}^i, s_{\text{cen}^*, t}, a)$ 表示为

$$p_{\text{cen}^*, t}(s_{\text{cen}^*, t}^i, s_{\text{cen}^*, t}, a) = \prod_{i=1}^N p_{\text{cen}^*, t}^i(s_{\text{cen}^*, t}^i, s_{\text{cen}^*, t}^i, a^i) \quad (5)$$

用 $\alpha_{\text{cen}^*, t}^i(L^i, L^i, a^i)$ 表示给定时隙 t 开始时刻节点 i 所有数据分组交付时间 $L_t^i = L^i$ 及个体行为 $a_t^i = a^i \in A_{\text{cen}^*, s}^i$ 时 $L_{t+1}^i = L^i$ 的条件概率; 用 $\beta_{\text{cen}^*, t}^i(e^i, e^i, a^i)$ 表示给定时隙 t 开始时刻节点 i 剩余能量 $e_t^i = e^i$ 及 $a_t^i = a^i \in A_{\text{cen}^*, s}^i$ 时 $e_{t+1}^i = e^i$ 的条件概率。由于给定 $a_t^i = a^i \in A_{\text{cen}^*, s}^i$ 时, 各节点所有数据分组交付时间和剩余能量互相独立, 式(5)可以进一步表示为

$$p_{\text{cen}^*, t}(s_{\text{cen}^*, t}^i, s_{\text{cen}^*, t}, a) = \prod_{i=1}^N \alpha_{\text{cen}^*, t}^i(L^i, L^i, a^i) \beta_{\text{cen}^*, t}^i(e^i, e^i, a^i) \quad (6)$$

首先, 推导 $\alpha_{\text{cen}^*, t}^i(L^i, L^i, a^i)$ 的取值。

1) 当 $L^i = 00 \cdots 0$ 时, 可知节点 i 在时隙 t 开始时刻数据队列为空。此情况下, 针对 $a^i \in A_{\text{cen}^*, s}^i = \{0\}$, 如果时隙 t 无数据分组到达节点 i , 则 $L^i = 00 \cdots 0$, 因此, $\alpha_{\text{cen}^*, t}^i(L^i, L^i, 0) = 1 - \lambda$; 反之, 则 $L^i = 0 \cdots 01$, 因此, $\alpha_{\text{cen}^*, t}^i(L^i, L^i, 0) = \lambda$; 而其他情况转移概率为 0。

2) 当 $L^i = 1f_2^i \cdots f_D^i$ 时, 可知时隙 t 的节点 i 队首数据分组无论在时隙 t 被发送与否均将在时隙 t 未被丢弃。此情况下, 对于任意 $a^i \in A_{\text{cen}^*, s}^i = \{0, 1\}$,

如果时隙 t 末无数据分组到达节点 i , 则 $L^i = f_2^i \cdots f_D^i 0$, 因此, $\alpha_{\text{cen}^*, t}^i(L^i, L^i, 0) = 1 - \lambda$; 反之, 则 $L^i = f_2^i \cdots f_D^i 1$, 因此, $\alpha_{\text{cen}^*, t}^i(L^i, L^i, 0) = \lambda$; 而其他情况转移概率均为 0。

3) 当 $L^i = 0f_2^i \cdots f_D^i$ 且非全 0 时, 可得知时隙 t 的节点 i 队首数据分组仅在发送时会被移出队列。不失一般性, 假设 L^i 自左向右第 k 位取值为 1, 且第 k 位之前的取值均为 0。此情况下, 对于任意 $a^i \in A_{\text{cen}^*, s}^i = \{0, 1\}$, 如果节点 i 在时隙 t 不发送数据分组且无数据分组到达时, 可得 $L^{i'} = f_2^i \cdots f_{k-1}^i 1 \cdots f_D^i 0$, 因此, $\alpha_{\text{cen}^*, t}^i(L^i, L^i, 0) = (1 - a^i)(1 - \lambda)$; 如果节点 i 在时隙 t 不发送数据分组且有数据分组到达时, 可得 $L^{i'} = f_2^i \cdots f_{k-1}^i 1 \cdots f_D^i 1$, 因此, $\alpha_{\text{cen}^*, t}^i(L^i, L^i, 0) = (1 - a^i)\lambda$; 如果节点 i 在时隙 t 发送数据分组且无数据分组到达时, 可得 $L^{i'} = f_2^i \cdots f_{k-1}^i 0 \cdots f_D^i 0$, 因此, $\alpha_{\text{cen}^*, t}^i(L^i, L^i, 0) = a^i(1 - \lambda)$; 如果节点 i 在时隙 t 发送数据分组且有数据分组到达时, 可得 $L^{i'} = f_2^i \cdots f_{k-1}^i 0 \cdots f_D^i 1$, 因此, $\alpha_{\text{cen}^*, t}^i(L^i, L^i, 0) = a^i\lambda$; 而其他情况转移概率均为 0。

其次, 针对以下两种情况推导 $\beta_{\text{cen}^*, t}^i(e^i, e^i, a^i)$ 的取值。

1) 当 $e^i < m$ 时, 可知节点 i 在时隙 t 开始时刻的能量不足以发送数据分组。此情况下, 对于 $a^i \in A_{\text{cen}^*, s}^i = \{0\}$, 如果节点 i 在时隙 t 末收集能量失败, 则 $e^{i'} = e^i$, 因此, $\beta_{\text{cen}^*, t}^i(e^i, e^i, 0) = 1 - \mu$; 反之 $e^{i'} = \min\{e^i + w, E\}$, 因此, $\beta_{\text{cen}^*, t}^i(\min\{e^i + w, E\}, e^i, 0) = \mu$; 而其他情况转移概率均为 0。

2) 当 $m \leq e^i \leq E$ 时, 可知节点 i 在时隙 t 开始时刻的能量足以用于发送数据分组。对于任意 $a^i \in A_{\text{cen}^*, s}^i = \{0, 1\}$, 如果节点 i 在时隙 t 不发送数据分组且收集能量失败, 则 $e^{i'} = e^i$, 因此, $\beta_{\text{cen}^*, t}^i(e^i, e^i, a^i) = (1 - \mu)(1 - a^i)$; 如果节点 i 在时隙 t 发送数据分组且在时隙 t 末收集能量失败, 则 $e^{i'} = e^i - m$, 因此, $\beta_{\text{cen}^*, t}^i(e^i - m, e^i, a^i) = (1 - \mu)a^i$; 如果节点 i 在时隙 t 不发送数据分组且成功收集能量, 则 $e^{i'} = \min\{e^i + w, E\}$, 因此 $\beta_{\text{cen}^*, t}^i(\min\{e^i + w, E\}, e^i, a^i) = \mu(1 - a^i)$; 如果节点 i 在时隙 t 发送数据分组且成功收集能量, 则 $e^{i'} = \min\{e^i - m + w, E\}$, 因此, $\beta_{\text{cen}^*, t}^i(\min\{e^i - m + w, E\}, e^i, a^i) = \mu a^i$; 而

其他情况下的转移概率均为 0。

决策规则及策略：考虑所有数据分组的交付时间，定义时隙 t 时的集中式决策规则为 $\pi_{\text{cen}^*,t}: S_{\text{cen}^*} \rightarrow A_{\text{cen}^*,s}$ 。此情况下的集中式策略 π_{cen^*} 可以表示为一系列集中式决策规则组成的序列，即 $\pi_{\text{cen}^*} \triangleq (\pi_{\text{cen}^*,1}, \pi_{\text{cen}^*,2}, \dots, \pi_{\text{cen}^*,t})$ 。标识考虑所有数据分组交付时间下所有可能集中式接入策略的集合为 Π_{cen^*} 。

奖励：标识时隙 t 的完整网络状态 $s_{\text{cen}^*,t} = s \in S_{\text{cen}^*}$ 及网络行为 $a_t = a \in A_{\text{cen}^*,s}$ 时的奖励为 $r_{\text{cen}^*,t}(s, a)$ ，并且定义为时隙 t 成功发送的数据分组数量的期望值，即

$$r_{\text{cen}^*,t}(s, a) \triangleq \sum_{i=1}^N a^i \prod_{j \neq i} (1 - a^j) \quad (7)$$

进一步，给定采用的集中式策略 π_{cen^*} 以及初始完整网络状态 $s_{\text{cen}^*,1} = s \in S_{\text{cen}^*}$ ，期望总奖励 $R^{\pi_{\text{cen}^*}}(s)$ 可定义为

$$R^{\pi_{\text{cen}^*}}(s) \triangleq E^{\pi_{\text{cen}^*}} \left[\sum_{t=1}^T r_{\text{cen}^*,t}(s_{\text{cen}^*,t}, a_t) \mid s_{\text{cen}^*,1} = s \right] \quad (8)$$

2.2 MDP 简化建模

考虑队首数据分组信息对决策制定的主导作用，本节对第 1.4 节所述仅考虑队首数据分组交付时间的集中式接入协议进行 MDP 简化建模，包括对其状态、行为、状态转移概率、决策规则、奖励以及期望总奖励的定义和推导。

简化状态：仅考虑队首数据分组交付时间，定义时隙 t 的简化网络状态 $s_{\text{cen},t}$ 为

$$s_{\text{cen},t} \triangleq (d_t^1, e_t^1, d_t^2, e_t^2, \dots, d_t^N, e_t^N) \quad (9)$$

$s_{\text{cen},t}^i \triangleq (d_t^i, e_t^i)$ 可视为时隙 t 时节点 i 的简化个体状态。通过列举所有可能的简化网络状态和简化个体状态可以分别得到简化网络状态空间 S_{cen} 和各节点 i 的简化个体状态空间 $S_{\text{cen},t}^i$ 。

行为：基于每个节点 i 在时隙 t 开始时刻的个体行为即接入概率 a_t^i ，定义时隙 t 的网络行为 a_t 为

$$a_t \triangleq (a_t^1, a_t^2, \dots, a_t^N) \quad (10)$$

同样地， $\sum_{i=1}^N a_t^i \leq 1$ ，并且当 $d_t^i > 0$ 及 $e_t^i \geq m$ 时， $a_t^i \in \{0, 1\}$ ，否则 $a_t^i = 0$ 。按此规则，可以得到简化网络状态值为 $s_{\text{cen},t}$ 时的网络行为空间 $A_{\text{cen},s}$ ，以及各节点 i 简化本地状态值为 $s_{\text{cen},t}^i$ 时的节点 i 个体

行为空间 A_{cen,s^i}^i 。

简化状态转移概率：用 $p_{\text{cen},t}(s_{\text{cen},t}^i, s_{\text{cen}}, a)$ 表示给定简化网络状态 $s_{\text{cen},t} = s_{\text{cen}} \in S_{\text{cen}}$ 及网络行为 $a_t = a \in A_{\text{cen},s}$ 时 $s_{\text{cen},t+1} = s_{\text{cen}}^i$ 的条件概率。用 $p_{\text{cen},t}^i(s_{\text{cen},t}^i, s_{\text{cen}}^i, a^i)$ 表示给定节点 i 简化个体状态 $s_{\text{cen},t}^i = s_{\text{cen}}^i \in S_{\text{cen}}^i$ 及行为 $a_t^i = a^i \in A_{\text{cen},s^i}^i$ 时 $s_{\text{cen},t+1}^i = s_{\text{cen}}^i \in S_{\text{cen}}^i$ 的条件概率。同样地，可以将 $p_{\text{cen},t}(s_{\text{cen},t}^i, s_{\text{cen}}, a)$ 表示为

$$p_{\text{cen},t}(s_{\text{cen},t}^i, s_{\text{cen}}, a) = \prod_{i=1}^N p_{\text{cen},t}^i(s_{\text{cen},t}^i, s_{\text{cen}}^i, a^i) \quad (11)$$

用 $\alpha_{\text{cen},t}^i(d^i, d^i, a^i)$ 表示给定时刻 t 开始时刻节点 i 队首数据分组交付时间 $d_t^i = d^i$ 及个体行为 $a_t^i = a^i \in A_{\text{cen},s^i}^i$ 时 $d_{t+1}^i = d^i$ 的条件概率；用 $\beta_{\text{cen},t}^i(e^i, e^i, a^i)$ 表示给定时刻 t 开始时刻节点 i 剩余能量 $e_t^i = e^i$ 及 $a_t^i = a^i \in A_{\text{cen},s^i}^i$ 时 $e_{t+1}^i = e^i$ 的条件概率。同样地，式(11)可以进一步表示为

$$p_{\text{cen},t}(s_{\text{cen},t}^i, s_{\text{cen}}, a) = \prod_{i=1}^N \alpha_{\text{cen},t}^i(d^i, d^i, a^i) \beta_{\text{cen},t}^i(e^i, e^i, a^i) \quad (12)$$

首先，针对以下 3 种情况推导 $\alpha_{\text{cen},t}^i(d^i, d^i, a^i)$ 的取值。

1) 当 $d^i = 0$ 时，可知节点 i 在时隙 t 开始时刻数据队列为空。此情况下 $a^i \in A_{\text{cen},s^i}^i = \{0\}$ ，如果时隙 t 末无数据分组到达节点 i ，则 $d^{i'} = 0$ ，因此， $\alpha_{\text{cen},t}^i(0, 0, 0) = 1 - \lambda$ ；反之，则 $d^{i'} = D$ ，因此， $\alpha_{\text{cen},t}^i(D, 0, 0) = \lambda$ ；而其他情况转移概率为 0。

2) 当 $d^i = 1$ 时，可知时隙 t 的节点 i 队首数据分组无论在时隙 t 被发送与否均将在时隙 t 末被丢弃。此情况下，对于任意 $a^i \in A_{\text{cen},s^i}^i = \{0, 1\}$ ，如果时隙 t 时节点 i 数据队列第二个数据分组比队首数据分组晚 $1 \leq l \leq D - 1$ 个时隙到达，可获得 $d^{i'} = l$ ，因此， $\alpha_{\text{cen},t}^i(l, 1, a^i) = \lambda(1 - \lambda)^{l-1}$ ；如果时隙 t 时节点 i 队首数据分组到达后 $D - 1$ 个时隙内均未有数据分组到达，可获得 $d^{i'} = 0$ ，因此 $\alpha_{\text{cen},t}^i(0, 1, a^i) = (1 - \lambda)^D$ ；而其他情况转移概率均为 0。

3) 当 $2 \leq d^i \leq D$ 时，可知时隙 t 的节点 i 队首数据分组仅在被发送时会被移出队列。此情况下，对于任意 $a^i \in A_{\text{cen},s^i}^i = \{0, 1\}$ ，如果节点 i 在时隙 t 不

发送数据时可获得 $d^i = d^i - 1$, 因此, $\alpha_{\text{cen},t}^i(d^i - 1, d^i, a^i) = 1 - a^i$; 如果节点 i 在时隙 t 发送数据且时隙 t 数据队列第二个数据分组比队首数据分组晚 $1 \leq l \leq D - 1$ 个时隙到达, 可获得 $d^i = l$, 因此, $\alpha_{\text{cen},t}^i(l, d^i, a^i) = a^i \lambda (1 - \lambda)^{l-d^i}$; 如果节点 i 在时隙 t 发送数据且时隙 t 队首数据分组到达后 $D - 1$ 个时隙内均未有数据分组到达, 可获得 $d^i = 0$, 因此 $\alpha_{\text{cen},t}^i(0, d^i, a^i) = a^i (1 - \lambda)^{D-d^i+1}$; 而其他情况转移概率均为 0。

其次, $\beta_{\text{cen},t}^i(e^i, e^i, a^i)$ 与第 2.1 节 $\beta_{\text{cen},t}^i(e^i, e^i, a^i)$ 的取值情况相同, 这里不再赘述。

决策规则及策略: 仅考虑队首数据分组交付时间, 定义时隙 t 时的集中式决策规则为 $\pi_{\text{cen},t}: S_{\text{cen}} \rightarrow A_{\text{cen},s}$ 。此情况下的集中式策略 π_{cen} 可以表示为 $\pi_{\text{cen}} \triangleq (\pi_{\text{cen},1}, \pi_{\text{cen},2}, \dots, \pi_{\text{cen},T})$ 。标识仅考虑队首数据分组交付时间下所有可能集中式接入策略的集合为 Π_{cen} 。

奖励: 标识时隙 t 的简化网络状态 $s_{\text{cen},t} = s \in S_{\text{cen}}$ 及网络行为 $a_t = a \in A_{\text{cen},s}$ 时的奖励为 $r_{\text{cen},t}(s, a)$, 并且定义为时隙 t 成功发送的数据分组数量的期望值, 即

$$r_{\text{cen},t}(s, a) \triangleq \sum_{i=1}^N a^i \prod_{j \neq i} (1 - a^j) \quad (13)$$

进一步, 给定采用的集中式策略 π_{cen} 以及初始简化网络状态 $s_{\text{cen},1} = s \in S_{\text{cen}}$, 期望总奖励 $R^{\pi_{\text{cen}}}(s)$ 可定义为

$$R^{\pi_{\text{cen}}}(s) \triangleq E^{\pi_{\text{cen}}} \left[\sum_{t=1}^T r_{\text{cen},t}(s_{\text{cen},t}, a_t) \middle| s_{\text{cen},1} = s \right] \quad (14)$$

2.3 最优集中式策略求解

尽管第 2.1 与第 2.2 节定义的状态及相关元素不同, 但均为有限时域 MDP 建模, 因此求解最优策略方法相同。鉴于此, 本节仅阐述第 2.2 节简化建模的最优集中式策略求解。

显而易见, 给定采用的集中式策略 π_{cen} 以及初始网络状态 $s_{\text{cen},1} = s$, 网络吞吐率 G 可以表示为 $R^{\pi_{\text{cen}}}(s) / T$ 。因此, 本节寻求使得任意初始简化网络状态 $s_{\text{cen},1} = s$ 下期望总奖励最大化的最优集中式策略 π_{cen}^* , 即寻求 $\pi_{\text{cen}}^* \in \Pi_{\text{cen}}$ 使得

$$R^{\pi_{\text{cen}}^*}(s) \geq R^{\pi_{\text{cen}}}(s), \quad s \in S_{\text{cen}}, \quad \pi_{\text{cen}} \in \Pi_{\text{cen}} \quad (15)$$

定义价值函数 $U_t^*(s)$ 为简化网络状态 $s_{\text{cen},t} = s \in S_{\text{cen}}$ 条件下从时隙 t 到时隙 T 的最大期望

总奖励。进一步, 考虑所有可能的下一时隙简化状态值 $s_{\text{cen},t+1} = s' \in S_{\text{cen}}$ 及其对应价值函数, 可以基于贝尔曼方程采用下述逆向归纳法^[28]求解 π_{cen}^* 。

算法 1 应用逆向归纳算法求解 π_{cen}^*

步骤 1 设置 $t=T$, 并计算

$$U_T^*(s) = \max_{a \in A_{\text{cen},s}} r_T(s, a), \quad \forall s \in S_{\text{cen}} \quad (16)$$

$$\pi_{\text{cen},T}^*(s) \in \arg \max_{a \in A_{\text{cen},s}} r_T(s, a), \quad \forall s \in S_{\text{cen}} \quad (17)$$

步骤 2 设置 $t=t-1$, 并计算

$$U_t^*(s) = \max_{a \in A_{\text{cen},s}} r_t(s, a) + \sum_{s' \in S_{\text{cen}}} p(s', s, a) U_{t+1}^*(s'), \quad \forall s \in S_{\text{cen}}, \quad (18)$$

$$\pi_{\text{cen},t}^*(s) \in \arg \max_{a \in A_{\text{cen},s}} r_t(s, a) + \quad (19)$$

$$\sum_{s' \in S_{\text{cen}}} p(s', s, a) U_{t+1}^*(s'), \quad \forall s \in S_{\text{cen}}$$

步骤 3 如果 $t=1$, 停止, 否则回到步骤 2。

3 分布式接入策略优化

根据第 1 节所述系统模型, 本节将对严格传输时限下能量收集 WSN 的分布式接入协议进行有限时域 Dec-MDP 建模, 并基于此建模求解 ε -最优分布式策略。受第 2.2 节启发, 本节同样将采用简化状态及相关定义。

3.1 Dec-MDP 简化建模

本节对第 1.4 节所述分布式接入协议进行 Dec-MDP 建模, 包括对其简化状态、行为、简化状态转移概率、简化占用状态、决策规则、奖励以及期望总奖励的定义和推导。

简化状态: 定义与第 2.2 节简化状态定义完全相同。

行为: 定义与第 2.2 节个体行为与网络行为定义相同, 但为了便于后续数值处理, 假设个体行为值仅可能为区间 $[0,1]$ 间隔为 Δa 的均匀采样。并且, 当 $d_i^i > 0$ 及 $e_i^i \geq m$ 时, $a_i^i \in (0,1]$; 否则 $a_i^i = 0$ 。按此规则, 可以得到简化网络状态值为 $s_{\text{cen},t}$ 时的网络行为空间 $A_{\text{dis},s}$, 以及各节点 i 简化本地状态值为 $s_{\text{cen},t}^i$ 时的节点 i 个体行为空间 A_{dis,s^i}^i 。

简化状态转移概率: 定义与第 2.2 节简化状态转移概率定义相同, 区别仅在于需要考虑离散化的网络行为空间 $A_{\text{dis},s}$ 和各节点 i 的个体行为空间 A_{dis,s^i}^i 。

决策规则及策略：定义时隙 t 时节点 i 的分布式决策规则为 $\pi_{dis,t}^i: S_{cen}^i \rightarrow A_{dis,t}^i$ 。因此，分布式策略 π_{dis} 可以表示为各节点一系列分布式决策规则组成的二维数组，即

$$\pi_{dis} \triangleq \begin{pmatrix} \pi_{dis,1}^1, \pi_{dis,2}^1, \dots, \pi_{dis,T}^1 \\ \pi_{dis,1}^2, \pi_{dis,2}^2, \dots, \pi_{dis,T}^2 \\ \vdots \\ \pi_{dis,1}^N, \pi_{dis,2}^N, \dots, \pi_{dis,T}^N \end{pmatrix} \quad (20)$$

标识所有可能分布式接入策略的集合为 Π_{dis} 。

另外，标识时隙 t 所有节点分布式决策规则构成的向量 $(\pi_{dis,t}^1, \pi_{dis,t}^2, \dots, \pi_{dis,t}^N)$ 为 $\pi_{dis,t}$ 。

简化占用状态及其更新：给定初始简化网络状态 $s_{cen,1}$ ，定义初始简化占用状态 η_1 为简化网络状态 $s_{cen,1}$ 取值的概率分布。定义时隙 t 的简化占用状态 η_t 为给定 η_1 及时隙 t 之前的所有分布式决策规则 $\pi_{dis,1}, \pi_{dis,2}, \dots, \pi_{dis,t-1}$ 时简化网络状态 $s_{cen,t} = s \in S_{cen}$ 取值的概率分布，即

$$\eta_t(s) \triangleq \Pr(s_{cen,t} = s | \eta_1, \pi_{dis,1}, \pi_{dis,2}, \dots, \pi_{dis,t-1}), s \in S_{cen} \quad (21)$$

因此，可根据更新函数 $\omega_t(\eta_t, \pi_{dis,t})$ 获得 $S_{cen,t+1} = s' \in S_{cen}$ 取值的概率分布 η_{t+1}

$$\eta_{t+1}(s') = \omega_t(\eta_t, \pi_{dis,t}) \triangleq \sum_{s \in S_{cen}} p_t(s', s, (\pi_{dis,t}^1(s_{cen,t}^1), \dots, \pi_{dis,t}^N(s_{cen,t}^N))) \cdot \eta_t(s), s' \in S_{cen} \quad (22)$$

奖励：标识时隙 t 在简化占用状态 η_t 及网络行为 $a_t = a \in A_{dis,t}$ 时的奖励为 $h_t(\eta_t, a)$ ，并且定义为时隙 t 成功发送的数据分组数量的期望值，即

$$h_t(\eta_t, a) \triangleq \sum_{s \in S_{cen}} \eta_t(s) r_{cen,t}(s, a) \quad (23)$$

进一步，给定采用的分布式策略 π_{dis} 以及初始简化网络状态 $s_{cen,1} = s \in S_{cen}$ ，期望总奖励 $Q^{\pi_{dis}}(s)$ 可定义为

$$Q^{\pi_{dis}}(s) \triangleq E^{\pi_{dis}} \left[\sum_{t=1}^T h_t(\eta_t, a_t) \middle| s_{cen,1} = s \right] \quad (24)$$

3.2 ε -最优分布式策略求解

显而易见，给定采用的分布式策略 π_{dis} 以及初始简化网络状态 $s_{cen,1} = s$ ，网络吞吐率 G 可以表示为 $Q^{\pi_{dis}}(s)/T$ 。因此，给定任意正数 ε ，本节寻求 ε -最优分布式策略 π_{dis}^ε ，即寻求 $\pi_{dis}^\varepsilon \in \Pi_{dis}$ 使得

$$Q^{\pi_{dis}^\varepsilon}(s) \geq Q^{\pi_{dis}}(s) - \varepsilon, s \in S_{cen}, \pi_{dis} \in \Pi_{dis} \quad (25)$$

标识给定 η_t 以及采用的分布式策略 π_{dis} 条件下从时隙 t 到时隙 T 的期望总奖励为 $V_t^{\pi_{dis}}(\eta_t)$ ，并基于贝尔曼方程可将其表示为

$$\begin{aligned} V_t^{\pi_{dis}}(\eta_t) &= h_t(\eta_t, \pi_{dis,t}) + V_{t+1}^{\pi_{dis}}(\omega_t(\eta_t, \pi_{dis,t})), \\ \forall t \in \{1, 2, \dots, T-1\} \\ V_T^{\pi_{dis}}(\eta_t) &= h_T(\eta_t, \pi_{dis,T}) \end{aligned} \quad (26)$$

基于此，本文应用马尔可夫策略搜索方法^[29-30]求解 π_{dis}^ε 。如图 2 所示，该算法的搜索树可以表示为一棵决策树，树的根为初始简化占用状态 η_1 ，不同树枝对应不同 $\pi_{dis,t}$ 及对应的立即奖励，树的节点对应不同 $\pi_{dis,t}$ 生成的 η_t 。树的节点旁所标记的一维向量对应了 $V_t^{\pi_{dis}}(\eta_t)$ 的上下界，可以通过树一直备份更新到根节点。

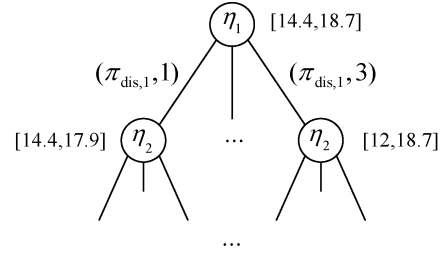


图 2 马尔可夫策略搜索的决策树

算法 2 应用马尔可夫策略搜索算法求解 π_{dis}^ε

主算法：

步骤 1 针对 $t=1, 2, \dots, T$ 以及任意 η_t ，使用简化建模下的最优集中式策略 π_{cen}^* 初始化 $V_t^{\pi_{dis}}(\eta_t)$ 的上界 $\bar{V}_t(\eta_t)$ ，即

$$\bar{V}_t(\eta_t) = \sum_{s \in S_{cen}} \eta_t(s) U_t^*(s) \quad (27)$$

然后使用不依赖于任何信息的随机产生的分布式策略 π_{rand} 初始化 $V_t^{\pi_{dis}}(\eta_t)$ 的下界 $V_t(\eta_t)$ ，即

$$V_t(\eta_t) = V_t^{\pi_{rand}}(\eta_t) \quad (28)$$

步骤 2 设置 $\tau=1$ ，若 $\bar{V}_\tau(\eta_\tau) - V_\tau(\eta_\tau) \leq \varepsilon$ ，则停止；否则，执行 $MPS_\tau(\eta_\tau)$ 算法后返回步骤 2。

$MPS_\tau(\eta_\tau)$ 算法：

步骤 3 若 $\bar{V}_\tau(\eta_\tau) - V_\tau(\eta_\tau) \leq \varepsilon$ ，则停止；否则，执行步骤 4~步骤 7 后返回步骤 3。

步骤 4 基于 $\bar{V}_\tau(\eta_\tau)$ 求解一组时隙 τ 时各节点贪婪分布式决策规则构成的 $\pi_{greedy, \tau} \triangleq (\pi_{greedy, \tau}^1,$

$\pi_{\text{greedy}, \tau}^2, \dots, \pi_{\text{greedy}, \tau}^N$), 即

$$\pi_{\text{greedy}, \tau}^N(s) \in \arg \max_{\pi_{\text{dis}, \tau} \in \Pi_{\text{dis}}} h_{\tau}(\eta_{\tau}, (\pi_{\text{dis}, \tau}^1(s_{\text{cen}, \tau}^1), \dots, \pi_{\text{dis}, \tau}^N(s_{\text{cen}, \tau}^N))) + \bar{V}_{\tau+1}(\omega_{\tau}(\eta_{\tau}, \pi_{\text{dis}, \tau})), \forall s \in S_{\text{cen}} \quad (29)$$

设 $\bar{V}_{T+1}(\omega_{T+1}(\eta_T, \pi_{\text{dis}, T}))$ 为 0。

步骤 5 从 $t = \tau$ 到 $t = 1$, 针对所有可能访问的 η_t 更新

$$\bar{V}_t(\eta_t) = h_t(\eta_t, (\pi_{\text{greedy}, t}^1(s_{\text{cen}, t}^1), \dots, \pi_{\text{greedy}, t}^N(s_{\text{cen}, t}^N))) + \bar{V}_{t+1}(\omega_t(\eta_t, \pi_{\text{greedy}, t})) \quad (30)$$

步骤 6 执行 $\text{MPS}_{\tau+1}(\omega_{\tau}(\eta_{\tau}, \pi_{\text{greedy}, \tau}))$ 算法。

步骤 7 如果 $V_t(\eta_t) < V_1^{\pi_{\text{greedy}}}$, 则针对 $t=1, 2, \dots, T$ 更新

$$V_t(\eta_t) = h_t(\eta_t, (\pi_{\text{greedy}, t}^1(s_{\text{cen}, t}^1), \dots, \pi_{\text{greedy}, t}^N(s_{\text{cen}, t}^N))) + V_{t+1}^{\pi_{\text{greedy}}}(\omega_t(\eta_t, \pi_{\text{greedy}, t})) \quad (31)$$

4 仿真结果

本节仿真对比完整建模下的最优集中式策略 π_{cen}^* 、简化建模下的最优集中式策略 π_{cen}^* 、简化建模下的 10^{-7} -最优分布式策略 $\pi_{\text{dis}}^{10^{-7}}$ 和最优静态接入策略的吞吐率性能。设置分布式接入建模的接入概率采样间隔 $\Delta a = 0.01$, 而最优静态策略要求各节点以最优固定接入概率发送数据。

策略计算复杂度及观测需求对比见表 1。前 3 种策略的计算复杂度均随用户数 N 呈指数增长, 但 $\pi_{\text{dis}}^{10^{-7}}$ 由于观测需求更低, 从而需要用更高的计算复杂度换取性能优化。另外, 最优静态接入策略复杂度最低, 但完全没有利用观测信息帮助决策, 因此可以预期将具有最差性能。

设置节点个数 $N=2$, 电池容量 $E=4$, 每次充电收集能量单元个数 $w=1$, 每次传输消耗能量单元个数 $m=1$, 初始能量 $e_1^1 = e_1^2 = 0$ 。图 3~图 7 每种情况下

的仿真结果均取 10^6 次仿真实验的均值。另外, 除非特殊说明, 仿真结果均为泊松能量到达下获得。

$\lambda = 0.3$ 、 $\mu = 0.5$ 、 $D=3$ 时不同时域长度下吞吐率性能对比如图 3 所示, 可以看出 π_{cen}^* 、 π_{cen}^* 的吞吐率性能几乎一致, 验证了 MDP 简化建模的有效性, 表明了队首数据分组交付时间相较于其他数据分组对决策制定起到决定性作用。同时, 可以发现 π_{cen}^* 、 π_{cen}^* 明显优于其他策略, 这是由于其利用每个时隙网络状态进行接入概率的动态规划, 因此可以完全避免节点间接入冲突且使得期望总奖励最大。此外, 可以发现 $\pi_{\text{dis}}^{10^{-7}}$ 的吞吐率性能明显优于最优静态策略。这是由于 $\pi_{\text{dis}}^{10^{-7}}$ 基于 $t=1$ 时的网络状态、后续时隙网络状态的演进以及个体状态制定了各节点的动态接入概率, 而静态最优接入策略仅根据网络参数设置最优固定接入概率。另外, 可以看出 $\pi_{\text{dis}}^{10^{-7}}$ 在泊松能量到达下仍然具有优异性能, 验证了简化考虑伯努利能量到达的合理性。

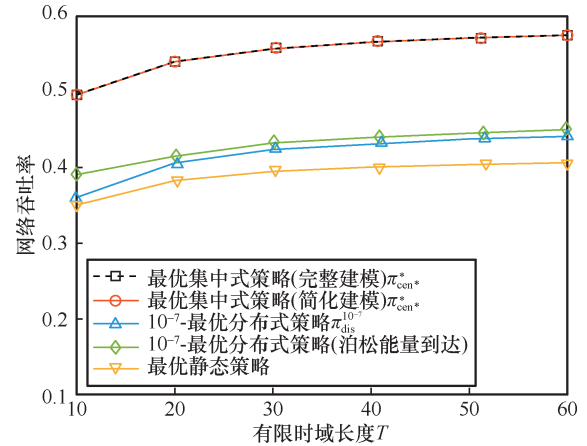


图3 $\lambda = 0.3$ 、 $\mu = 0.5$ 、 $D=3$ 时不同时域长度下吞吐率性能对比

$T=30$ 、 $\lambda = 0.3$ 、 $D=3$ 时不同能量收集成功率下吞吐率对比如图 4 所示, 可以看出 π_{cen}^* 、 π_{cen}^* 在所有情况下仍然性能一致且表现最优, 而 $\pi_{\text{dis}}^{10^{-7}}$ 次之。此外, 所有策略的网络吞吐率均逐渐随着 μ 的增大而改善, 但增速趋于缓慢。这是因为 μ 增大

表 1

策略计算复杂度及观测需求对比

| 算法 | 复杂度 | 观测需求 |
|-------------------|-------------------------------------|---------------------|
| 最优集中式策略(完整建模) | $O([2^D(E+1)]^{2N} \cdot 2^N)$ | 每个时隙均为完全观测 |
| 最优集中式策略(简化建模) | $O([(D+1)(E+1)]^{2N} \cdot 2^N)$ | 每个时隙均为完全观测 |
| 10^{-7} 最优分布式策略 | $O((1/\Delta a + 1)^{(D+1)(E+1)N})$ | 仅第一时隙为完全观测其余时隙为部分观测 |
| 最优静态策略 | $O(1)$ | 无须观测 |

使各节点剩余能量提高,使得因能量不够而无法接入信道的情况更少发生;而又受限于各节点能量队列有限长度, μ 增大到一定程度时,各节点的剩余能量将趋于饱和,因此不会进一步改善吞吐率。可以再次发现 $\pi_{dis}^{10^{-7}}$ 在泊松能量到达下仍然具有优异性能。

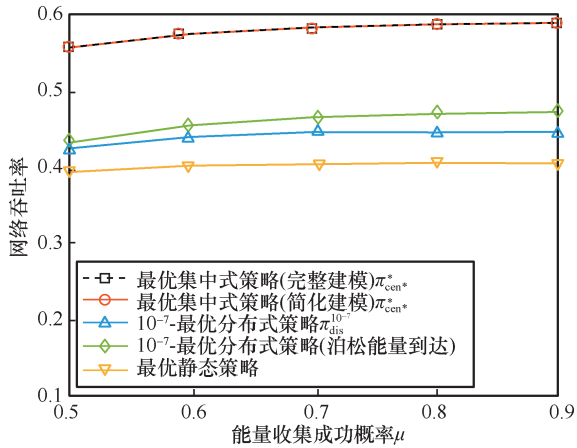


图 4 $T=30$ 、 $\lambda=0.3$ 、 $D=3$ 时不同能量收集成功率下吞吐率对比

$T=30$ 、 $\mu=0.5$ 、 $D=3$ 时不同数据分组到达率下吞吐率对比如图 5 所示,可以看出 π_{cen}^* 、 π_{cen}^* 在所有情况下性能仍然一致且表现最优,而 $\pi_{dis}^{10^{-7}}$ 次之,再次验证了本文策略设计的有效性。此外,随着 λ 的增大,网络吞吐率逐渐增大,但增速趋于缓慢。这是由于当 λ 增大到一定程度时,信道拥挤使得数据分组更频繁地由于交付期限到期被丢弃,因此无法继续提高网络的吞吐率性能。

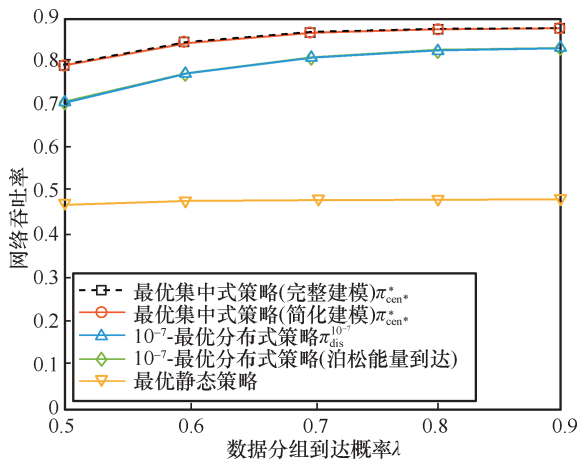


图 5 $T=30$ 、 $\mu=0.5$ 、 $D=3$ 时不同数据分组到达率下吞吐率对比

$T=30$ 、 $\lambda=0.7$ 、 $\mu=0.5$ 时不同传输时限下吞吐率对比如图 6 所示,可以看出 π_{cen}^* 、 π_{cen}^* 在所有

情况下性能仍然一致且表现最优,而 $\pi_{dis}^{10^{-7}}$ 次之。此外,随着 D 的增大,网络吞吐率逐渐增大,但增速趋于缓慢。其原因是 D 的增大使得数据分组的传输紧迫度降低,从而使得数据分组由于交付期限到期被丢弃的情况减少,但当 D 增大到一定程度时,网络吞吐率主要受数据分组到达率影响,因此吞吐率变化趋于平缓。同时,可以发现 π_{cen}^* 、 π_{cen}^* 与 $\pi_{dis}^{10^{-7}}$ 之间的吞吐率差距随着 D 增加逐渐减小,可以归因为传输紧迫度的降低减少了决策所依据信息不完整的负面作用。可以再次发现 $\pi_{dis}^{10^{-7}}$ 在泊松能量到达下仍然具有优异性能。

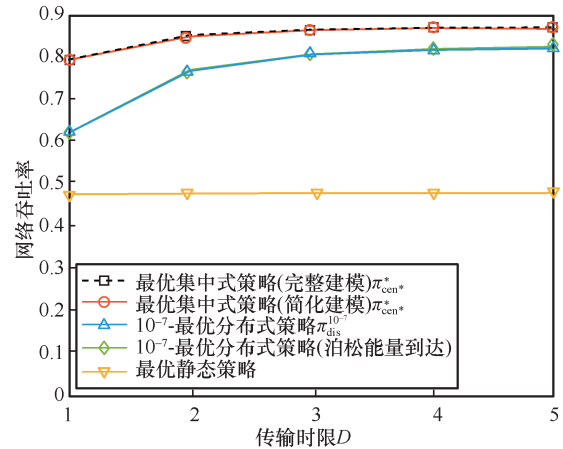


图 6 $T=30$ 、 $\lambda=0.7$ 、 $\mu=0.5$ 时不同传输时限下吞吐率对比

最后,鉴于本文建模和优化可以很容易拓展至不同节点不同数据到达率的情况, $T=30$ 、 $D=3$ 、 $\mu=0.8$ 、节点 1 数据分组到达率为 0.5 时,节点 2 在不同数据分组到达率情况下,两节点吞吐率性能及平均剩余能量对比如图 7 所示,对比了 10^{-7} -最优分布式策略 $\pi_{dis}^{10^{-7}}$ 各节点的网络吞吐率和平均剩余能量水平。可以发现,随着节点 2 数据分组到达概率的增大,节点 2 吞吐率性能较节点 1 有显著提升,而由于更频繁传输,节点 2 平均剩余能量水平相较于节点 1 逐渐降低,表明了能量充足时, $\pi_{dis}^{10^{-7}}$ 使得各节点仅会基于传输紧迫性的优先级选择相应接入概率,而剩余能量的区别对决策不会造成影响。

5 结束语

本文设计了传输时限约束下能量收集 WSN 的集中式和分布式接入协议,允许各节点根据传输紧迫性和能量储备程度采用不同接入概率。以最大化网络吞吐率为优化目标,使用 MDP 对集中式接入分

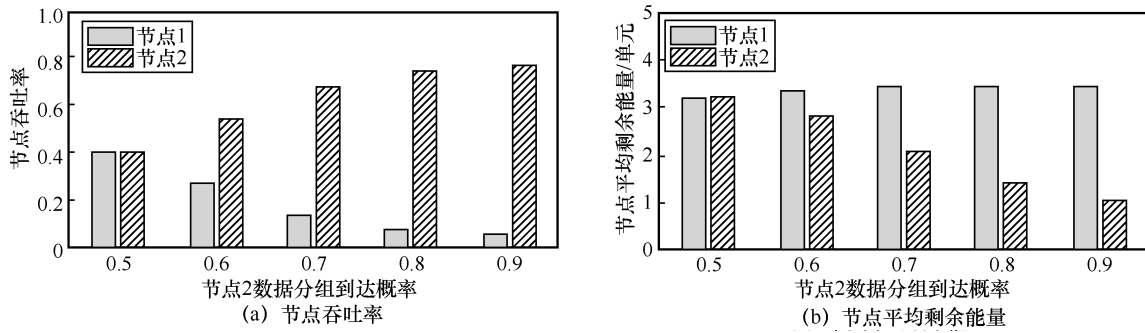


图7 $T=30$ 、 $D=3$ 、 $\mu=0.8$ 、节点1数据分组到达率为0.5时, 节点2在不同数据分组到达率下, 两节点吞吐量性能及平均剩余能量对比

别进行了完整建模和简化建模, 并求解了两种建模的最优集中式策略; 然后使用 Dec-MDP 对分布式接入协议进行了简化建模和 ϵ -最优分布式策略求解。本文采用的简化建模思想仅基于各节点队首数据分组 (而不是所有数据分组) 交付时间推导各节点传输紧迫性的变化, 有效降低了建模和优化复杂度。仿真结果验证了简化建模的有效性, 表明了队首数据分组交付时间对决策起到决定性作用, 并验证了所提出策略相较于其他策略的性能优势。本文的建模和优化方法适用于任意独立同分布能量到达过程, 仿真结果也表明基于伯努利能量到达所提出策略在其他能量到达过程下仍然具有优异性能。另外, 通过加入内嵌马尔可夫模型对能量收集过程进行建模, 本文的建模和优化方法能够扩展至非独立同分布能量到达过程。尽管如此, 结合具体网络的能量收集过程进行理论分析是非常有必要的, 我们将在未来工作中通过真实场景实验对能量收集过程的统计特性进行刻画, 以增加本文理论工作的应用价值。

本文聚焦于一跳 WSN, 所提策略尽管可以基于分簇拓扑^[23]应用于大型多跳 WSN, 但簇头数量选取、簇头之间通信和簇内通信可以根据传输时限和能量收集约束进行进一步联合优化, 这将作为未来工作之一。

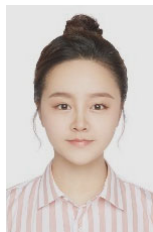
参考文献:

- [1] QUINTERO V L, ESTEVEZ C, ORCHARD M E, et al. Improvements of energy-efficient techniques in WSNs: a MAC-protocol approach[J]. IEEE Communications Surveys & Tutorials, 2018, 21(2): 1188-1208.
- [2] HASSAN G, HASSANEIN H S. MoT: a deterministic latency MAC protocol for mission-critical IoT applications[C]//Proceedings of 2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC). Piscataway: IEEE Press, 2018: 588-593.
- [3] FARAG H, GIDLUND M, ÖSTERBERG P. A delay-bounded MAC protocol for mission- and time-critical applications in industrial wireless sensor networks[J]. IEEE Sensors Journal, 2018, 18(6): 2607-2616.
- [4] THOMAS D, SHANKARAN R, ORGUN M, et al. Energy-efficient military surveillance: coverage meets connectivity[J]. IEEE Sensors Journal, 2019, 19(10): 3902-3911.
- [5] LUVISOTTO M, PANG Z B, DZUNG D. High-performance wireless networks for industrial control applications: new targets and feasibility[J]. Proceedings of the IEEE, 2019, 107(6): 1074-1093.
- [6] 王哲, 李陶深, 叶进, 等. 基于不确定理论的能量收集可靠性建模及规划[J]. 通信学报, 2018, 39(5): 166-176.
WANG Z, LI T S, YE J, et al. Reliability modeling and planning of energy harvesting based on uncertainty theory[J]. Journal on Communications, 2018, 39(5): 166-176.
- [7] LU X, WANG P, NIYATO D, et al. Wireless networks with RF energy harvesting: a contemporary survey[J]. IEEE Communications Surveys & Tutorials, 2014, 17(2): 757-789.
- [8] BAE Y H. Analysis of optimal random access for broadcasting with deadline in cognitive radio networks[J]. IEEE Communications Letters, 2013, 17(3): 573-575.
- [9] ZHANG Y J, GUAN F Y, LO Y H. Coding-based slotted ALOHA for broadcasting multislot messages with delivery deadline[J]. IEEE Transactions on Vehicular Technology, 2018, 67(8): 7882-7886.
- [10] ZHANG Y J, LO Y H, SHU F, et al. Achieving maximum reliability in deadline-constrained random access with multiple-packet reception[J]. IEEE Transactions on Vehicular Technology, 2019, 68(6): 5997-6008.
- [11] BAE Y H. Optimal retransmission-based broadcasting under delivery deadline constraint[J]. IEEE Communications Letters, 2015, 19(6): 1041-1044.
- [12] DENG L, DENG J, CHEN P N, et al. On the asymptotic performance of delay-constrained slotted ALOHA[C]//Proceedings of 2018 27th International Conference on Computer Communication and Networks (ICCCN). Piscataway: IEEE Press, 2018: 1-8.
- [13] ELSAWY H. Characterizing IoT networks with asynchronous time-sensitive periodic traffic[J]. IEEE Wireless Communications Letters, 2020, 9(10): 1696-1700.
- [14] ZHAO L L, CHI X F, QIAN L, et al. Analysis on latency-bounded reliability for adaptive grant-free access with multipackets reception (MPR)

in URLLCs[J]. IEEE Communications Letters, 2019, 23(5): 892-895.

- [15] JU H, ZHANG R. Throughput maximization in wireless powered communication networks[J]. IEEE Transactions on Wireless Communications, 2014, 13(1): 418-428.
- [16] BIASON A, ZORZI M. Battery-powered devices in WPCNs[J]. IEEE Transactions on Communications, 2017, 65(1): 216-229.
- [17] MICHELUSI N, ZORZI M. Optimal adaptive random multiaccess in energy harvesting wireless sensor networks[J]. IEEE Transactions on Communications, 2015, 63(4): 1355-1372.
- [18] BIASON A, DEY S, ZORZI M. A decentralized optimization framework for energy harvesting devices[J]. IEEE Transactions on Mobile Computing, 2018, 17(11): 2483-2496.
- [19] DIBANGOYE J S, AMATO C, BUFFET O, et al. Optimally solving dec-POMDPs as continuous-state MDPs[J]. Journal of Artificial Intelligence Research, 2016, 55: 443-497.
- [20] BAE Y H. Modeling timely-delivery ratio of slotted aloha with energy harvesting[J]. IEEE Communications Letters, 2017, 21(8): 1823-1826.
- [21] BAE Y H, BAEK J W. Performance analysis of delay-constrained traffic in a cognitive radio network with RF energy harvesting[J]. IEEE Communications Letters, 2019, 23(12): 2177-2181.
- [22] NIYATO D, WANG P. Delay-limited communications of mobile node with wireless energy harvesting: performance analysis and optimization[J]. IEEE Transactions on Vehicular Technology, 2014, 63(4): 1870-1885.
- [23] JIA D Y, ZHU H H, ZOU S X, et al. Dynamic cluster head selection method for wireless sensor network[J]. IEEE Sensors Journal, 2016, 16(8): 2746-2754.
- [24] ULUKUS S, YENER A, ERKIP E, et al. Energy harvesting wireless communications: a review of recent advances[J]. IEEE Journal on Selected Areas in Communications, 2015, 33(3): 360-381.
- [25] SHAVIV D, ÖZGÜR A. Universally near-optimal online power control for energy harvesting nodes[J]. IEEE Journal on Selected Areas in Communications, 2016, 34(12): 3620-3631.
- [26] BLASCO P, GUNDUZ D, DOHLER M. A learning theoretic approach to energy harvesting communication system optimization[J]. IEEE Transactions on Wireless Communications, 2013, 12(4): 1872-1882.
- [27] WANG X, GONG J, HU C, et al. Optimal power allocation on discrete energy harvesting model[J]. EURASIP Journal on Wireless Communications and Networking, 2015, 2015(1): 1-14.
- [28] BAXTER L A. Markov decision processes: discrete stochastic dynamic programming[J]. Technometrics, 1995, 37(3): 353.
- [29] DIBANGOYE J S, AMATO C, DONIEC A. Scaling up decentralized MDPs through heuristic search[C]//Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence. 2012: 217-226.
- [30] DIBANGOYE J S, AMATO C, DONIEC A, et al. Producing efficient error-bounded solutions for transition independent decentralized MDPS[C]//AAMAS '13: Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems. 2013: 539-546.

[作者简介]



杨澳钦（1999-），女，南京理工大学硕士生，主要研究方向为无线网络协议设计和优化。



宫傲宇（1997-），男，洛桑联邦理工学院硕士生，主要研究方向为无线网络协议设计及优化。



房婷（1998-），女，南京理工大学硕士生，主要研究方向为无线网络协议设计和优化。



邓磊（1989-），男，博士，深圳大学助理教授，主要研究方向为时延受限网络通信。



李强（1973-），男，博士，鹏城实验室正高级工程师，主要研究方向为物联网与5G/B5G。



张一晋（1982-），男，博士，南京理工大学教授，主要研究方向为序列设计、无线网络与人工智能。