

Learning-based Autonomous Channel Access in the Presence of Hidden Terminals

摘要

我们考虑自主信道接入 (AutoCA) 问题, 其中一组终端尝试以分布式方式通过公共无线信道发现与接入点 (AP) 的通信策略。由于不规则的拓扑结构和终端的有限通信范围, AutoCA 面临的实际挑战是隐藏终端问题, 该问题在无线网络中因降低吞吐量和延迟性能而著名。为了应对这一挑战, 本文提出了一种新的多智能体深度强化学习范式, 称为 MADRL-HT, 专为存在隐藏终端的 AutoCA 量身定制。MADRL-HT 利用拓扑洞察力, 将每个终端的观察空间转换为独立于终端数量的可扩展形式。为了补偿部分可观测性, 我们提出了一种回溯机制, 使得终端可以根据载波感知的信道状态以及来自 AP 的反馈来推断其隐藏终端的行为。提出了基于窗口的全局奖励函数, 指示终端最大化系统吞吐量, 同时在学习过程中平衡终端的传输机会。考虑到短数据包的机器类型通信, 大量的数值实验验证了我们的解决方案的卓越性能, 以传统的载波侦听多路访问和冲突避免 (CSMA/CA) 协议为基准。

关键词: 多智能体深度强化学习; Wi-Fi 隐藏终端; 近端策略优化

1 引言

自主信道接入 (AutoCA) 是一类多路访问控制 (MAC) 问题, 其中一组终端以分散的方式在共享信道上与接入点 (AP) 进行通信 [8]-[3]。由于多址干扰, 任何时候只有一个终端可以与 AP 通信; 当两个或多个终端同时传输时, 会发生冲突并且传输失败。在 AutoCA 中, 终端的目标是自主发现一组传输策略, 使得系统吞吐量最大化并保证终端之间的公平性。AutoCA 对于缺乏良好管理基础设施的通信网络至关重要, 例如在免许可频段上运行的通信系统 [4]-[1], 以提高高带宽效率和低延迟服务。

深度强化学习多路访问 (DRLMA) 的最新进展为解决 AutoCA 问题开辟了新的视角。借助 DRLMA [8] [4] [1] [6] [7] [2] [5], 具有智能的分布式终端学会合作 (而不是竞争) 来和谐地共享频谱。特别是, 终端组需要适应彼此的传输策略, 以避免冲突并平衡信道访问机会。为此, AutoCA 问题被表述为马尔可夫决策过程 (MDP) 或部分可观察 MDP (POMDP), 取决于终端是否完全可观察到环境。然后通过 DRL 或多代理 DRL (MADRL) 来解决 MDP 或 POMDP, 其中终端从其过去的传输历史记录、其他终端的载波感知行为以及从 AP/信道反馈的附加信息中逐步学习一组良好的策略。与传统的随机访问策略相比, DRLMA 在吞吐量、延迟和抖动方面通常表现出更好的性能, 展示了在各种网络场景中解决 AutoCA 问题的巨大潜力

无线网络中一个非常麻烦的问题是隐藏终端问题。隐藏终端是指超出彼此通信范围的终端。隐藏终端无法通过载波侦听来检测彼此的传输行为，因此它们到公共接收器的传输可能会发生冲突，从而导致频谱浪费。例如，在 Wi-Fi 网络中，当存在隐藏终端时，LBT 协议就会崩溃，原因是不准确的载波感应信道状态；并且吞吐量和延迟性能显著恶化。为了缓解这个问题，IEEE 802.11 标准引入了请求发送/清除发送 (RTS/CTS) 机制。然而，RTS/CTS 并不是理想的协议，因为它会产生额外的握手开销和新问题，例如隐藏接收器和暴露接收器因此，它甚至会降低 CSMA/CA 的性能。这在有效负载较小和最新 Wi-Fi 标准的机器类型通信中更为明显，其中带宽变得越来越大，产生持续时间越来越短的数据包。另一方面，先前关于 DRLMA 的工作仅关注理想情况，其中所有终端都可以相互听到，并且完全忽略了隐藏终端问题。在隐藏终端的实际设置中，终端可能观察到不准确的信道状态；它的传输可能会与它从未见过的隐藏终端的传输发生冲突。更糟糕的是，隐藏终端的传输策略随时间变化，给终端带来非平稳的隐藏动态。在这种背景下，设计 DRLMA 算法来解决隐藏终端的影响是一项重要且具有挑战性的任务。

2 相关工作

IEEE 802.11 中针对 AutoCA 问题的标准解决方案是具有二进制指数退避 (BEB) 的 CSMA/CA。具体来说，CSMA/CA 的退避时间是从 $[0, CW]$ 中随机采样的，其中 CW 是竞争窗口的长度。当发生冲突时， CW 会加倍，以减少争用的概率。 CW 的设置对 CSMA/CA 的效率有很大影响。由于 BEB 是一种启发式方案，最近的一些工作采用 DRL 来学习在不同网络条件下设置 CW 的更好策略，从而优化系统吞吐量。

与优化 CSMA/CA 之上的参数相反，一种更有前景、更有趣的方法可以充分利用 DRL 的潜力，那就是从头开始学习 AutoCA 协议。在这一领域的工作中，早期研究侧重于异构设置，其中智能终端（或具有中央控制器的终端集群）学习与使用传统 MAC 协议操作的其他终端共存，例如 Aloha、TDMA、CSMA/CA 和 LTE。这些工作提出了单智能体强化学习问题，其中智能体从其观察中学习适应其他非智能体的未知策略。与 MARL 相比，单智能体 RL 的挑战性较小，因为智能体面临的环境是固定的。

通过载波侦听，代理可以完全观察到信道占用情况。在具有隐藏终端的更实际的设置中，终端可以观察到不准确的信道状态；它的传输可能会与它从未见过的隐藏终端的传输发生冲突。更糟糕的是，隐藏终端的传输策略随时间变化，给每个终端带来非平稳的隐藏动态。在这个在上下文中，设计 DRLMA 算法来解决隐藏终端的影响是一项具有挑战性的任务，我们的目标是在本文中解决这个问题。

3 本文方法

我们考虑传统的 Wi-Fi 基本服务集 (BSS)，其中一个 AP 和 N 个终端形成星形拓扑，如图 1 (a) 所示。终端的索引为 $n = 1, 2, \dots, N$ 。我们重点关注从终端到 AP 的上行传输。所有终端共享相同的上行频谱，称为公共信道。时间被划分为多个时段。我们考虑饱和流量设置，即终端的传输缓冲区在任何间隙都非空，并且只要有空闲间隙，终端就有数据包要传输。在整篇论文中，我们假设传输失败只能由冲突引起，而不是底层物理信道的质量。

并非所有终端都在彼此的通信范围内。例如，在图 1 (a) 中，终端 A 和 B 在彼此的通信范围内。因此，当 A 空闲时，终端 A 可以感知终端 B 的行为（即是否发送），反之亦然。另一方面，终端 A 和 C 超出了彼此的通信范围，因此它们无法从载波侦听中推断出彼此的传输——它们对彼此来说是“隐藏终端”。在存在隐藏终端的情况下，本文旨在解决 AutoCA 问题，即一组代理如何自主地达成多路访问协议，以最大化网络吞吐量，同时保证彼此之间的公平性。更严格地，我们使用下面的公平性度量来制定 AutoCA 问题。

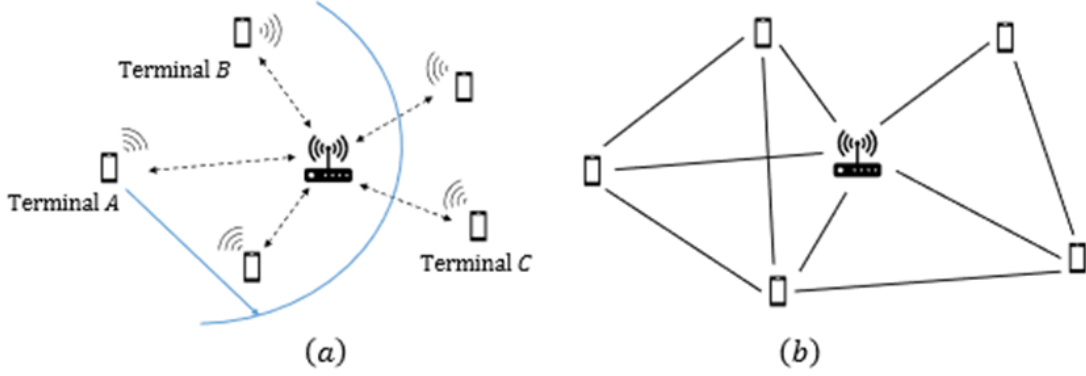


图 1. WiFi-BSS

3.1 多智能体强化学习建模

一般的 MARL 问题可以定义为一个元组 (N, S, O_n, A, r_n) ，其中 $N=1, 2, \dots, N$ 表示 N 个代理的集合； S 表示智能体所在环境的状态空间； O_n 表示第 n 个智能体的观察空间； A 表示每个智能体的动作空间； $r_n : S \times A \times S \rightarrow \mathbb{R}$ 表示第 n 个智能体的即时奖励函数，它定义了环境分配给智能体的奖励。在上述文中，与 AutoCA 相关的 MARL 问题如图 2 所示。在任意时间步 t ，第 n 个终端的动作定义为 $a_n^t \in A = \{transmit, idle\}$ ；环境状态 $S_t \in S$ 代表信道使用情况，是所有终端动作的集合。每个终端都有一个 S_t 的本地观测值，用 O_n^t 表示。具体地， O_n^t 由第 n 个终端的载波侦听信道状态和来自 AP 的 ACK 反馈确定。根据观察 O_n^t 和收到的奖励，每个代理采取行动 a_n^t 。动作集 a_n^t 引导环境移动到新的状态 S_{t+1} ，并且奖励 r_n 被分配给每个终端以评估其最后的动作。在 MARL 的过程中，终端旨在以分布式方式发现一组传输策略-每个终端从与其他终端的交互、其部分观察到的信道状态以及它收到的奖励中学习-以避免冲突并最大化 BSS 吞吐量，同时保证彼此之间的公平性。如图 2

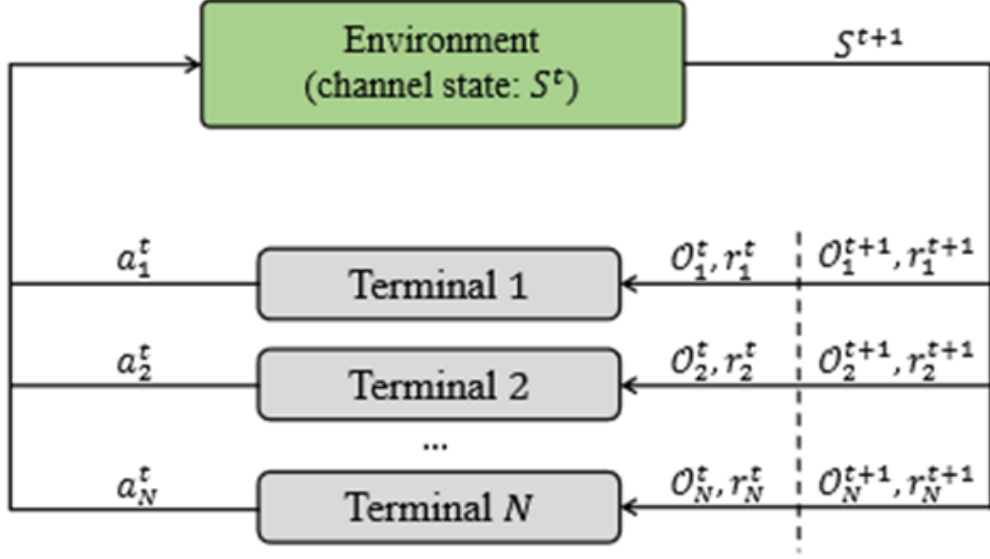


图 2. MARL

3.2 终端聚类

对于 Wi-Fi BSS 中的任意终端来说，它所面临的环境由其他 $N-1$ 个终端组成。为了制定与其他终端良好共存的良好传输策略，第一步是推断其他每个终端的传输策略。然而，我们问题的部分可观察性使得推理成为一项难以捉摸的任务，因为不可能从有限的载波感知信道状态和 AP 的反馈中推断出其他 $N-1$ 个终端的传输历史。

考虑图上的两个顶点 A 和 B。当且仅当顶点 A 可以在至少 1 跳中到达顶点 B 时，顶点 B 是顶点 A 的 1 跳邻居。

给定定义 1，由于 BSS 的星形拓扑，我们从图 1(b) 中得到了两个直接结果。

在 Wi-Fi BSS 中，所有终端都是 AP 的一跳 (OH) 邻居。

对于 BSS 中的终端 A 来说，任何其他终端要么是终端 A 的 OH 邻居，要么是终端 A 的两跳 (TH) 邻居。

综上所述，对于 WiFi BSS 中的第 n 个终端，我们可以将其他 $N-1$ 个终端分为两组：一组由第 n 个终端的所有 OH 邻居组成，记为 $OH(n)$ ，另一个由第 n 个终端的所有 TH 个邻居组成，记为 $TH(n)$ 。图 3 给出了一个示例，其中 BSS 中的任何其他终端是终端 A 或 C 的 OH 或 TH 邻居。我们强调不同的终端具有不同的 OH 和 TH 邻居集合。

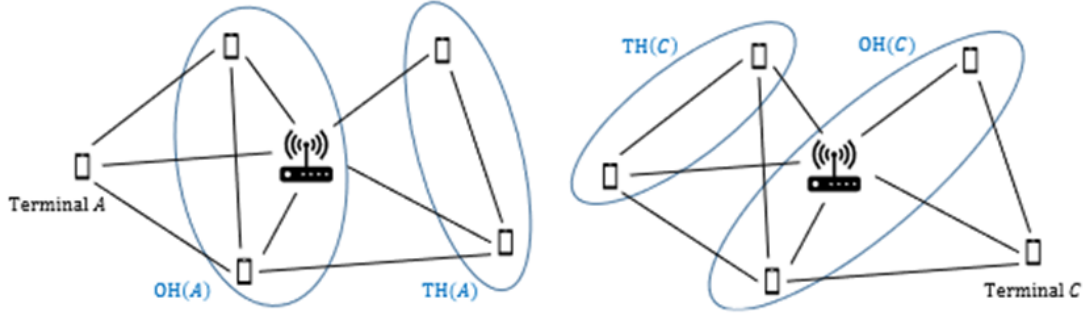


图 3. 聚类

另一个关键观察结果是，对于第 n 个终端， $OH(n)$ 中的所有终端都可以被视为单个合并终端，因为 1) 第 n 个终端以相同的方式感知它们；2) 他们的行为（传输或不传输）对第 n 个终端具有相同的效果。同样， $TH(n)$ 中的所有终端也可以被视为单个合并终端。因此，从第 n 个终端的角度来看，所有其他终端的传输动作可以概括为两个合并终端的传输动作： $OH(n)$ 和 $TH(n)$ 。

3.3 网络设置

由于每个终端的传输历史是一个时间序列，因此我们将参与者和批评者设计为循环神经网络（RNN），并以双向 LSTM 作为循环单元。如图 4 所示，参与者和评论家共享相同的架构，唯一的区别是输入和输出。

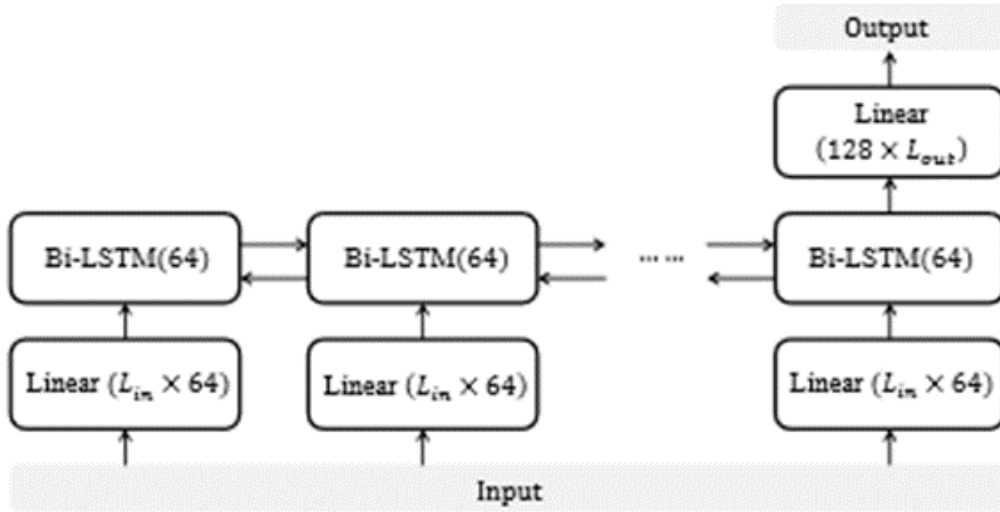


图 4. RNN

网络的输入对于第 n 个终端的参与者，输入是一个 $3 \times W$ 维矩阵，由第 n 个终端的传输历史、 $OH(n)$ 和回溯窗口中的 $TH(n)$ 组成，对于批评者来说，输入是一个 $N \times W$ 维矩阵，由回溯窗口中所有终端的传输历史组成。深度神经网络第一层是一个由 64 个神经元组成的线性层，其中对于 actor 来说 $L_{in}=3$ ，对于 critic 来说 $L_{in}=N$ 。第二层是具有 64 个隐藏神经元

的双向 LSTM 层。第三层是一个由 128 个神经元组成的线性层，其中演员的 $L_{out}=2$ ，评论家的 $L_{out}=1$ 。ReLU 激活函数应用于每个线性层的输出

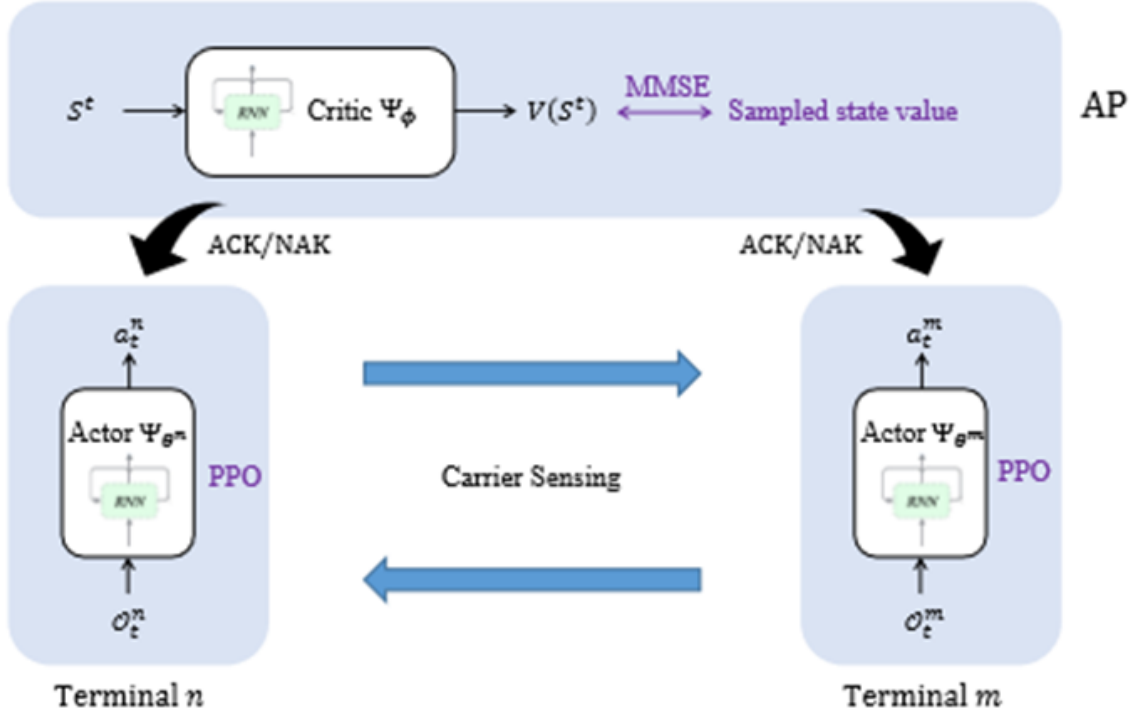


图 5. MADRL-HT

4 复现细节

4.1 与已有开源代码对比

本次对全文进行了复现，其中关键的就是对于此问题 MDP 模型的建立，将其复现为代码，与 CSMA 协议。本文并没有开源，仅在文中给出了算法的伪代码形式，因此复现过程对于 PPO 与 CSMA 查阅了已有代码的形式，经过调试和修改后，算法可以收敛，能接近论文水平。

4.2 实验环境搭建

采用的是 pytorch 的模型框架，主要实验环境如下：matplotlib numpy tqdm torch=1.11.1

4.3 创新点

本次实验的创新点相对较少，主要是对现有工作的复现。创新之处在于修改了模型的细节参数，对模型的学习效果有一定提升。同时，由于模型细节在文中并没有很好的介绍，加入残差网络有效防止过拟合。并设置不同的网络拓扑结构，来进行实验。

5 实验结果分析

对于实验，设置了如下几种拓扑模型，来进行 $\{A|B\}$, $\{AB|BC\}$ 并分别对其在 MADRL-HT 与 CSMA/CA 环境进行实验。用此符号”|” 隔开，意味着两个终端为 TH 邻居。

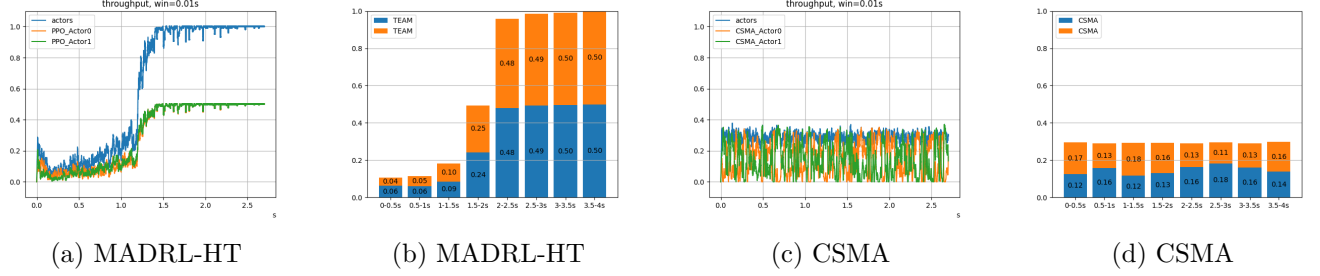


图 6. Topo1

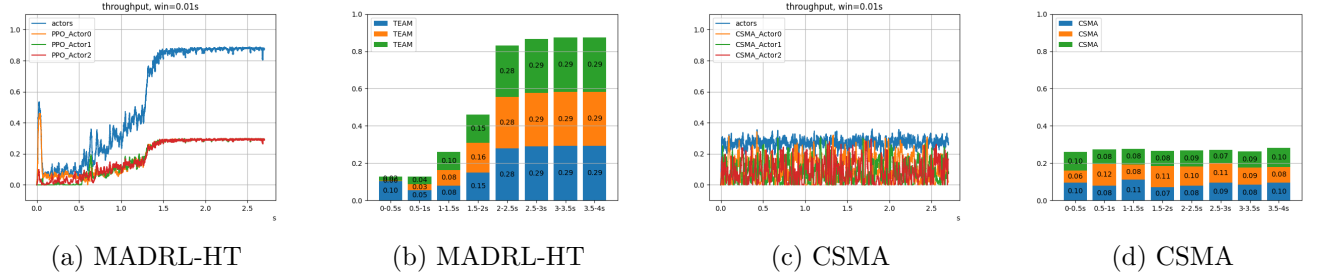


图 7. Topo2

我们首先研究各个终端所达到的吞吐量, 对于 CSMA/CA-BSS, 每个终端的信道访问策略是固定的。各个终端的吞吐量具有波动的特点——各个终端的吞吐量呈现出脉冲特性, 在 Topo1 和 Topo2 中都是相当不公平的。隐藏终端降低了 BSS 吞吐量并加剧了终端之间的不公平性。

对于 MADRL-HT-BSS, 各个终端实现的吞吐量几乎相等, 并且在 Topo1 和 Topo2 中 BSS 吞吐量均优于 CSMA/CA-BSS。一个有趣的观察是, MADRL-HT 在 Topo2 中比在 Topo1 中学习得更好: 收敛速度更快, 并且实现了最佳 BSS 吞吐量。

6 总结与展望

本次实验的不足之处主要在复现的算法相对固定, 缺少创新性的改进。同时从复现结果来看, 离论文结果尚有一定差距, 主要是模型未调参以及研究方向尚处于起步阶段, 对细节不清晰导致的。但是在具有隐藏终端的 AutoCA 问题中, MADRL-HT 表现出突出特点, 具有很大的进步空间。

参考文献

- [1] Abhishek Kumar, Gunjan Verma, Chirag Rao, Ananthram Swami, and Santiago Segarra. Adaptive contention window design using deep q-learning. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4950–4954, 2021.
- [2] Oshri Naparstek and Kobi Cohen. Deep multi-user reinforcement learning for distributed dynamic spectrum access. *IEEE Transactions on Wireless Communications*, 18(1):310–323, 2019.
- [3] Yulin Shao, Qi Cao, Soung Chang Liew, and He Chen. Partially observable minimum-age scheduling: The greedy policy. *IEEE Transactions on Communications*, 70(1):404–418, 2022.
- [4] X. Yang and N. Vaidya. On physical carrier sensing in wireless ad hoc networks. In *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, volume 4, pages 2525–2535 vol. 4, 2005.
- [5] Kok-Lim Alvin Yau, Peter Komisarczuk, and D. Teal Paul. Enhancing network performance in distributed cognitive radio networks using single-agent and multi-agent reinforcement learning. In *IEEE Local Computer Network Conference*, pages 152–159, 2010.
- [6] Yiding Yu, Soung Chang Liew, and Taotao Wang. Non-uniform time-step deep q-network for carrier-sense multiple access in heterogeneous wireless networks. *IEEE Transactions on Mobile Computing*, 20(9):2848–2861, 2021.
- [7] Yiding Yu, Soung Chang Liew, and Taotao Wang. Multi-agent deep reinforcement learning multiple access for heterogeneous wireless networks with imperfect channels. *IEEE Transactions on Mobile Computing*, 21(10):3718–3730, 2022.
- [8] Eustathia Ziouva and Theodore Antonakopoulos. Cdma/ca performance under high traffic conditions: throughput and delay analysis. *Computer Communications*, 25(3):313–321, 2002.