鲁棒多智能体系统验证

Raphaël Berthon 亚琛工业大学 德国亚琛 巴黎-萨克雷大学、法国国家科学研究中心、巴黎-萨克雷高等师范学院 法国吉夫叙尔伊维特 rberthon@lmf.cnrs.fr

Joost-Pieter Katoen 亚琛工业大学 德国亚琛 katoen@cs.rwth-aachen.de

Munyque Mittelmann 巴黎北方大学计算机科学实验室、法国国家科学研究中心、索邦巴黎北方大学 法国维勒塔讷兹 mittelmann@lipn.univ-paris13.fr

Aniello Murano 那不勒斯费德里科二世大学 意大利那不勒斯 aniello.murano@unina.it

arXiv:2602.13405v1 [cs.LO] 13 Feb 2026

摘要

随机多智能体系统是自主控制器、通信协议和网络物理基础设施的核心建模框架。然而，在许多此类系统中，转移概率仅根据数据估算，因此可能部分未知或受到扰动。本文研究不完美信息下随机多智能体系统中鲁棒策略的验证。在这类系统中，联盟必须在应对不确定系统转移、部分观测和对抗性智能体的同时，满足时序规范。通过聚焦有界记忆策略，我们针对交替时间时序逻辑的一种概率型、基于观测的扩展，提出了鲁棒模型检测问题。我们刻画了该问题在不同扰动概念下的复杂度，从而阐明鲁棒性在随机多智能体验证中的计算成本，并支持在不确定环境中使用有界记忆策略。

关键词

模型检测、多智能体系统、策略推理

1 引言

基于人工智能的技术已广泛应用于各个领域，包括机器人学、自主智能体和计算机视觉。这些技术的普及凸显了确保计算系统可信性、正确性和鲁棒性的挑战。形式化验证为应对这些挑战提供了有原则的方法。特别是，模型检测[19] 提供了全自动化技术：将系统表示为标记迁移模型，并用时序逻辑公式描述其正确行为。

多智能体系统（MAS）由多个自主组件构成，并在共享环境中交互 [51]。多智能体系统的演化由智能体的行为决定。交替时间时序逻辑（ATL） [2] 是一种重要的形式化方法，可用于推理多智能体系统中的策略。除传统的时序模态外，ATL还包含策略模态，用以表达智能体联盟如何协作或竞争以实现其目标。由于随机事件和智能体不可预测的行为，多智能体系统的若干方面天然具有不确定性。这种不确定性可以通过实验或过往观测进行量化，并由随机模型刻画，例如马尔可夫决策过程（MDP）和随机多智能体系统。概率ATL（PATL） [17] 将ATL扩展到概率设定，使人们能够推理智能体在具有随机转移的系统中交互时的随机化策略能力。另一方面，多智能体系统中的不确定性也可能源于智能体对系统的部分可观测性。在这种不完美信息设定下，即使对于确定性多智能体系统，具有完美记忆的策略能力的模型检测也是不可判定的 [24], , 因此有必要审慎选择受限但仍具表达力的策略类别 [8, 12]。

尽管可以测量随机事件发生的频率，但其确切概率往往未知，系统也可能面临扰动。例如，不同的气象中心经常会对同一地区和时间给出不同的降水概率。另一个例子是基于模型的强化学习，其中智能体根据数据估计智能体-环境交互模型（例如，MDP [38]）。由于模型是通过与环境交互学习得到的，其转移容易出现误差。在这类情形下，策略行为难以应对，因为它既需要处理系统转移中的不确定性，又要与其他可能合作或对抗的智能体互动。

本文研究鲁棒策略的验证方法，即研究如何确保策略在转移概率受到扰动的情况下仍满足时序规范。我们还考虑是否存在满足类 PATL 规范的联盟策略。我们不再假设联合动作之后的转移概率固定，而是考虑概率范围处于给定扰动界限内的随机多智能体系统。由此得到的设定同时涵盖联盟策略不确定性以及确切转移概率的不确定性。

据我们所知，这是首篇研究具有 ATL 目标且转移概率不固定的随机多智能体系统的工作。此前，包括在完全信息的情形下，也未曾探索这一方向。我们考虑了三种鲁棒性定义。第一种采用 ε-扰动，遵循经典设定：随机系统中的每个转移概率都可以偏离标称值 [16] 至多 ε。另外两种定义则更为复杂，基于参数化系统，其中转移概率表示为有限个参数上的有理函数。我们研究参数数量固定的情形

这是与 AAMAS 2026 上发表的同名论文的扩展版本，其中包含证明细节的技术附录。

以及参数数量任意（无界）的情形。这一区分至关重要，因为参数数量任意的情形通常会导致复杂度显著跃升。

在我们的设定中，智能体的策略取决于它们对系统的观测，这体现了不完美信息。尽管近期研究已证明，在不完美信息和无记忆策略条件下，PATL 的模型检测问题是可判定的 [6, 7], ，我们也允许使用有界记忆策略。我们考虑基于自动机的策略，这类策略广泛应用于验证 [5] ，并且相较于 ATL 中考虑的标准有界记忆策略，表达能力更强。我们在命题 4.2 中对这一改进进行了形式化说明，证明基于自动机的策略严格推广了经典的 k-回忆策略。与此同时，这种额外的表达能力并不会增加我们设定下模型检测的最坏情况复杂度。

从实践角度来看，这一设定很有意义，因为完全记忆并非总是可行的。事实上，在经济学、机器人学或人工智能等现实应用中，智能体往往受限于计算资源、处理能力和时间限制。与系统交互的人类也可能面临类似限制，只能保留并处理有限的信息量 [29, 30]。基于观测的有界记忆策略能够在处理不确定性的同时保证计算上的可行性，并避免过度简化（无记忆策略）或不切实际（无界记忆）这两个极端。在部分观测条件下，采用更丰富的记忆表示尤为重要，因为有界记忆是实现可判定性的少数途径之一 [14]。因此，有界记忆在灵活性和效率之间取得了平衡，这对实际场景十分重要。

	普遍可达性（记忆型）	基于观测的 PATL（有界记忆）
ε-扰动。	在 **P** （定理 7.1*） [16]	见 $\Sigma_2^{\mathbf{P}}$ (定理 7.6**)
固定参数。	属于 **NP类 ∩ co-NP类** (定理 7.2*)	见 $\Sigma_2^{\mathbf{P}}$ (定理 7.8**)
无界	见 $\forall\mathbb{R}$(定理 7.3*)	见 $\Sigma_3^{\mathbb{R}}$ （定理 7.9*）

表 1：结果概要。标有 * 的结果允许采用随机策略。标有 ** 的结果考虑联盟采用确定性策略、联盟之外允许采用随机策略的情形。

贡献。 本文给出了在不确定随机多智能体系统中获取有界记忆鲁棒策略的复杂度结果。我们表明，我们对有界记忆的定义比回忆访问过的最近 b 个状态严格更一般。我们考虑逻辑 PATL，用于推理具有有界记忆的基于观测的策略。随后，我们定义了一种模型检测问题变体，允许存在潜在扰动，并确定了 PATL 规范的复杂度。我们的方法有两大优势。一方面，有界记忆是在计算成本与信息保留之间的一种权衡。另一方面，我们可以保证策略能够抵御模型受到的（有界）扰动。我们给出了具有合理复杂度的新模型检测结果，涵盖从“简单”的可达性目标到描述联盟行为的更复杂目标。虽然我们的可达性结果（定理 7.1、7.2、7.3）与现有文献中的结果较为接近，但它们是我们的 PATL 结果（定理 7.6、7.8、7.9）的重要构件。后面这些结果及其证明都是全新的，在技术上也颇具挑战，因为它们需要结合多智能体系统与参数化模型检测。

我们不使用最近访问的 n个状态来表示策略的记忆，而是使用包含 n 个状态的自动机，从而增强策略能力。虽然随机策略的表达能力更强，但对于确定性策略，我们得到了更好的复杂度结果。我们考虑不同类型的扰动：用参数表示的扰动，这些参数可以为不同动作所共用，因而能够对依赖关系建模；以及两个子情形：**(i)** 参数数量固定；**(ii)** 可能的扰动可取区间 ε 内的任意值。表 1 总结了我们的主要结果。我们主要关注成员资格结果。即便在基本情形下，一些困难性结果也可能难以获得。特别是，具有无界参数的普遍可达性，是存在可达性的对偶问题；众所周知，后者的 $\exists\mathbb{R}$-困难性仍是一个开放问题 [33]。进一步研究 $\Sigma_3^{\mathbb{R}}$ 时，实数层级中的困难性结果尤其难以获得 [43]。我们还在附录中讨论了不同方法为何无法给出 $\Sigma_2^{\mathbf{P}}$-困难性结果。对于参数数量固定的情形，目前尚无已知的**NP** ∩ **co-NP** -困难问题；若存在这样一个困难问题，则意味着 **P** ≠ **NP**。

2 相关工作

我们的工作与多智能体系统概率逻辑、区间和参数化马尔可夫决策过程（MDP），以及鲁棒策略合成方面的研究相关。

随机多智能体系统的验证。 针对以概率逻辑给出的规范验证随机多智能体系统，已得到广泛研究，其中包括使用概率交替时序逻辑（PATL）[17], 概率交替时间 μ演算 [45] 以及概率策略逻辑（PSL） [3]进行规范描述。并发随机博弈的验证也已通过 PRISM 模型检测器实现[36, 37]。近期研究探讨了在不完美信息条件下，针对 PATL 的模型检测问题，并考察了被验证联盟采用确定型 [6] 和混合策略 [7]的情形，其中策略为无记忆策略。最后， [11]研究了采用概率自然策略的 PATL 变体。自然策略也能表示有界记忆，但若直接应用本文提出的方法，就需要将其表示为自动机，导致表示规模呈指数级增长。

区间马尔可夫决策过程。区间马尔可夫链（IMC）通过区间值转移概率推广了马尔可夫链：对于任何转移集合包含在该区间内的系统，策略都必须成立。与我们的方法类似，区间马尔可夫链是一种用于对转移概率确切值存在不确定性的概率系统进行建模的工具 [26]。在区间马尔可夫链上对线性时序逻辑（LTL）公式进行模型检测属于**指数空间**，且是**PSPACE**困难的 [9]。推广到区间马尔可夫决策过程（IMDP）后，已有研究证明，使用椭球法对概率计算树逻辑（PCTL）进行模型检测可在多项式时间内完成 [16] 。然而，这种方法难以直接推广到基于观测的策略。研究表明，在区间马尔可夫决策过程中，为多个目标寻找鲁棒随机策略是**PSPACE**困难的 [26]。另一种针对不确定马尔可夫决策过程中的 PCTL 性质进行验证的方法

将凸不确定集视为区间的一种推广 [40]。类似地， [15] 研究了每个转移都具有不确定性集合且目标为平均收益目标的马尔可夫决策过程，并给出了NP∩co-NP 成员资格结果。

参数化马尔可夫决策过程。 一种对不确定性和鲁棒性进行一般化的方法是考虑参数。具体而言，转移概率可以表示为关于给定参数集的方程。区间马尔可夫链考虑转移概率的区间，而参数化马尔可夫链的转移概率则由关于固定实值参数集的有理数系数多项式给出。这种表示推广了区间马尔可夫链，因为转移概率之间可能存在依赖关系。文献 [4] 研究了规范采用概率计算树逻辑的参数化马尔可夫链模型检测问题。该研究表明，在单变量情形下，存在量化 PCTL 模型检测问题是**NP类**完全问题，并找出了可在多项式时间内解决模型检测问题的 PCTL 片段。相比之下，对于参数化马尔可夫决策过程，即使是最简单的情形也颇具挑战性 [33]：当参数数量固定时，任何可达性目标都可以在**NP类**中解决；但当参数数量任意时，大多数问题都变为 $\exists\mathbb{R}$完全问题。我们对固定与无界扰动的区分正是源于这些结果。另一个相关问题是，为参数化马尔可夫决策过程中未指定的参数计算一种实例化，使所得的马尔可夫决策过程满足给定的时序逻辑规范。文献 [21] 使用凸优化技术研究了这一问题。

一般而言，在参数化马尔可夫决策过程中区分满足区域与违反区域的问题很难，因为这些区域由非线性函数界定 [32]。参数化马尔可夫决策过程已被用于在双人部分可观测随机博弈框架中综合策略 [1]。这种框架将部分可观测马尔可夫决策过程（POMDP）推广到两个智能体，其中只有一个智能体能够完全观测系统。该综合方法将问题化简为一系列参数化马尔可夫决策过程的综合问题，并结合混合整数线性规划求解。

鲁棒策略综合。文献 [47]研究了针对具有可达性和期望成本目标的不确定部分可观测马尔可夫决策过程综合鲁棒无记忆策略的问题。与区间马尔可夫决策过程类似，不确定部分可观测马尔可夫决策过程考虑一组概率区间。所提出的解决方案采用凸优化，作者证明，对于无记忆行为策略，该问题是NP困难的。此项工作进一步扩展到有限状态策略，同时保留相同的复杂度下界 [22]。针对具有椭球不确定性的马尔可夫决策过程综合鲁棒策略的问题，已结合 PCTL 规范开展研究，并应用于计算能源定价中的风险限制策略 [41]。实验评估显示，s-矩形鲁棒马尔可夫决策过程取得了令人鼓舞的结果 [35, 50], ，但要将这些技术应用于多智能体系统，目前仍为时尚早。

人们也从其他角度研究了多智能体系统的鲁棒性。鲁棒交替时间逻辑 [39] 是 ATL 的扩展，用于表达在纯确定性设定下，智能体联盟容忍环境假设被违反的能力。环境假设被描述为一种时序逻辑性质。违反意味着在系统执行的某些时刻，该假设的真值可能为“假”。最近，这一概念被扩展到了概率设定 [54]。我们研究模型扰动下的验证，而 Yu 等人 [52] 研究了智能体的动作可能以一定概率失效的情形。

3 预备知识

本文固定有限非空的智能体集合 Ag、动作集合 Ac 和原子命题 AP。我们用 $\boldsymbol{c}$ 表示动作元组 $(c_a)_{a\in \mathrm{Ag}}$，其中每个智能体各有一个动作，这类元组称为动作配置。给定动作配置 $\boldsymbol{c}$ 和 $C \subseteq \mathrm{Ag}$，我们令 c_C 表示 C中的智能体所采取的动作，并令 $\boldsymbol{c}_{-C}$为 $(c_b)_{b\notin C}$。类似地，我们令 $\mathrm{Ag}_{-C} = \mathrm{Ag} \setminus C$。我们用 $\mathbb{N}_+$ 表示集合 {1、2、.....}。

分布。 设 X 为一个有限非空集合。 X上的（概率）分布是一个函数 $\mathrm{d}: X \to [0,1]$，满足 $\sum_{x\in X}\mathrm{d}(x) = 1$。令 $\mathrm{Dist}(X)$ 表示 X上的分布集合。对于 $\mathrm{d}(x)$，我们将 $x \in \mathrm{d}$ 写作 > 0。若对于某个元素 $x \in X$，有 $\mathrm{d}(x) = 1$，则 d 是点分布（亦称狄拉克分布）。若对于 $i \in I$，d_i是 X_i上的一个分布，则记作 $X = \prod_{i\in I} X_i$，d_i的乘积分布是由 $\mathrm{d}: X \to [0,1]$ 定义的分布，其满足 $\mathrm{d}(x) = \prod_{i\in I}\mathrm{d}_i(x_i)$。

随机系统。随机多智能体系统（或简称系统）G 是一个元组 $(\mathrm{St}, \mathrm{L}, \delta, v, (\mathrm{obs}_a)_{a\in \mathrm{Ag}})$，其中：(i) St 是一个有限且非空的状态集；(ii) $\mathrm{L}: \mathrm{St} \times \mathrm{Ag} \to 2^{\mathrm{Ac}} \setminus \{\emptyset\}$ 是一个合法性函数，用于定义每个智能体在每个状态下可用的动作；我们用 $\mathrm{L}(\boldsymbol{s})$ 表示元组 $(\mathrm{L}(s,a))_{a\in \mathrm{Ag}}$；(iii) 对于每个状态 $s \in \mathrm{St}$ 和每个动作 $\boldsymbol{c} \in \mathrm{L}(\boldsymbol{s})$，随机转移函数 δ 给出转移概率 $\delta(s,\boldsymbol{c})(s')$：当每个玩家 $a \in \mathrm{Ag}$ 执行动作 c_a时，状态会从 s转移到任意 $s' \in \mathrm{St}$，并且注意 $\delta(s,\boldsymbol{c}) \in \mathrm{Dist}(\mathrm{St})$；(iv) v: $\mathrm{St} \to 2^{\mathrm{AP}}$ 是一个标记函数。(v) 对每个 $a \in \mathrm{Ag}$，obs_a 是一个观测函数 $\mathrm{obs}: \mathrm{St} \to O_a$，其观测集为 O_a。

对于每个状态 $s \in \mathrm{St}$ 和联合动作 $\boldsymbol{c} \in \prod_{a\in \mathrm{Ag}} \mathrm{L}(s,a)$，我们假设存在一个状态 $s' \in \mathrm{St}$，使得 $\delta(s,\boldsymbol{c})(s')$ 非零。也就是说，每个状态都能通过合法动作到达一个后继状态，因此 $\boldsymbol{c} \in \mathrm{L}(s,a)$。

例 *3.1*（*River*）。考虑一个系统 G_{river}，其中两家公司共用一条河流。每个步骤中，每家公司都有两个可选动作：将废水直接排入河中（动作 d），或先进行处理再排入河中（动作 t）。原子命题表示河流水质是否已达到低（命题 low）或高（命题 $high$）水平。该系统如图 1 所示。命题 low和 $high$分别仅在状态 q_{low}和 q_{high}中为真。

在状态 q_{normal} 中，没有命题为真，这表示水质处于正常水平。如果两家公司都排放废水，则水质必然下降（从高到正常，再从正常到低）。同样，如果两家公司都处理水，则水质会提升（从低到正常，再从正常到高）。如果一家公司处理水，而另一家公司排放废水，结果则不是确定型的。如果水质原本处于正常水平，则水质可能以 50% 的概率升高或降低。如果水质处于高水平（低水平的情况类似），则水质可能以 75% 的概率保持不变，或以 25% 的概率降低（分别为升高）。

马尔可夫模型。 部分可观测马尔可夫决策过程（POMDP）是一种单玩家随机系统，为每个智能体都指定了一个观测函数。当所有观测函数

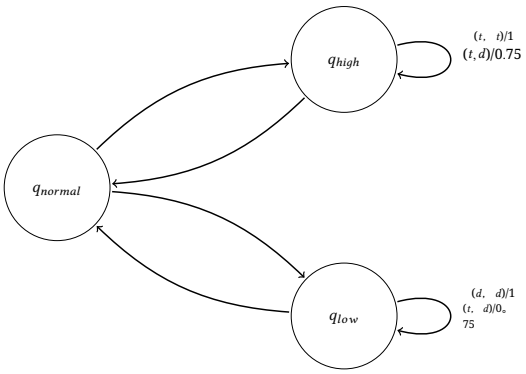

图 1：系统 $\mathcal{G}_{river}$ 表示两家公司之间交互的系统。转移函数以箭头上的标签表示。每次从状态 q 转移到 q'的转移，都标注了一个或多个形如$(x, y)/z$的标签，其中(x, y) 表示公司在状态q下的联合动作，而 z 表示到达状态 q'的概率。联合动作 (d, t) 的转移概率与(t, d)相同，因此省略。

都是单射时，该系统是完全可观测的，此时我们得到一个马尔可夫决策过程 （MDP），其中省略了观测。

马尔可夫链是一种马尔可夫决策过程，对每个 $s \in \mathrm{St}$ 都有 $|\mathrm{L}(s)| = 1$。我们通常将马尔可夫链 M 表示为一个元组 $(\mathrm{St}\ p)$，其中 St 是一个可数的非空状态集合，且 $p \in \mathrm{St} \to \mathrm{Dist}(\mathrm{St})$是转移矩阵。

运行。 系统 G 中的博弈路径 或路径，是一个状态的无限序列 $\pi = s_0 s_1 \cdots \in \mathrm{St}^\omega$，并且存在一个联合动作序列 $\boldsymbol{c}_0 \boldsymbol{c}_1 \cdots$，使得 $\boldsymbol{c}_i \in \mathrm{L}(s_i)$ 且 $s_{i+1} \in \delta(s_i, \boldsymbol{c}_i)$（即 $\delta(s_i, \boldsymbol{c}_i)(s_{i+1}) > 0$）对每个 $i \geq 0$都成立。我们用 π_i 表示 s_i，用 $\pi_{\geq i}$ 表示从位置 i开始的 π 后缀。有限路径称为历史，所有历史的集合记为 Hist。我们用 $\mathrm{last}(h)$表示历史 h 的最后一个状态，用 $len(h)$表示 h的大小。我们将观测函数 $\mathrm{obs} : \mathrm{St} \to O_a$扩展到路径，得到 $\mathrm{obs} : \mathrm{St}^\omega \to O_a^\omega$。

4 一般策略

接下来，我们将介绍一类一般的基于观测的策略，涵盖无限记忆和有界记忆两种情况。

定义 *4.1*（一般有界记忆策略）。 设 $\mathrm{G} = (\mathrm{St}, \mathrm{L}, \delta, v, (\mathrm{obs}_a)_{a \in \mathrm{Ag}})$ 是一个随机系统，其动作集合为 Ac，观测集合为 O。在最一般的情况下，智能体 a的策略是无界函数 $\sigma : \mathrm{obs}_a(\mathrm{Hist}) \to \mathrm{d}(\mathrm{Ac})$。

我们用Str_a 表示智能体 a的无界策略集合。我们采用 [5] 的定义 10.97，并按照 [20] 的做法扩展该定义，以纳入随机化和部分观测。如果策略 σ 可以表示为一个调度器 $(Q, act, \Delta, start)$，其中 Q是一个可能无限的模式（或记忆状态）集合， $\Delta : Q \times O \to \mathrm{Dist}(Q)$ 是一个随机转移函数， $act : Q \times O \to \mathrm{Dist}(\mathrm{Ac})$ 随机选择下一步动作，而 $start : O \to \mathrm{Dist}(Q)$ 根据观测 o 随机选择一个初始模式，则称该策略为有界回忆的一般随机策略。与 [20], 一样，如果 $b = \infty$，则策略具有无限记忆；如果 $|Q| = b$，则具有有限记忆 $b \in \mathbb{N}_+$；如果 $|Q| = 1$，则为无记忆策略，它

如果 act 是随机化的，但其他所有分布均为单元素集合，则称其为行为型策略；如果 $start$ 是随机化的，但其他所有分布均为单元素集合，则称其为混合型策略；如果所有分布均为单元素集合，则称其为纯型（或确定型）策略。如果 obs 不是单射，我们称该策略为基于观测的策略。我们在附录中证明，这一定义涵盖了其他定义，例如 [6, 7, 12]中的定义。我们对 b-有界回溯策略的定义使用有限状态调度器。正如我们接下来将展示的，这比回忆最近的 b 个状态更一般（见命题 4.2）。

首先，我们指出相较于 Belardinelli 等人 [6, 7]所采用的方法，我们在表示不完美信息多智能体系统方面存在的差异。我们将具有不完美信息的随机系统定义为 $\mathrm{G} = (\mathrm{St}, \mathrm{L}, \delta, v, (\mathrm{obs}_a)_{a \in \mathrm{Ag}})$，而他们则考虑 $\mathrm{G} = (\mathrm{St}, \mathrm{L}, \delta, v, (\sim_a)_{a \in \mathrm{Ag}})$，其中 $\sim_a$是 $\mathrm{St} \times \mathrm{St}$ 上的一个等价关系。这两个不完美信息定义是等价的，因为对于 $s, s' \in \mathrm{St}$，我们可以令 $\mathrm{obs}_a(s) = \mathrm{obs}_a(s')$ 当且仅当 $s \sim_a s'$。然而，我们的定义使得使用调度器定义基于观测的策略更加简单，尤其是函数 $\Delta : Q \times O \to \mathrm{Dist}(Q)$、 $act : Q \times O \to \mathrm{Dist}(\mathrm{Ac})$ 和 $start : O \to \mathrm{Dist}(Q)$。

我们对有限记忆策略的定义不同于策略推理逻辑中的通常定义，例如 [12],，后者考虑 k-顺序策略，定义为 $\sigma : \mathrm{obs}_a(O^k) \to \mathrm{d}(\mathrm{Ac})$，其中 $k \in \mathbb{N}$。我们的基于调度器的有限记忆策略严格地比顺序策略更一般，形式化地说：

命题 4.2。 以下结论成立：

- 对于每个 k-顺序策略，都存在一个基于调度器的有限记忆策略，在相同历史下执行相同的动作。
- 存在一个马尔可夫决策过程 $\mathcal{G}_\vartheta$ 和一个线性时序逻辑 公式φ_ϑ，使得某个基于调度器的有限记忆策略满足φ_ϑ 在$\mathcal{G}_\vartheta$ 上的概率为1，但无论 k如何，都不存在满足该条件的 k-顺序策略。

证明概要。对于第二点，考虑图 2 中的系统 G_ϑ 和公式 $\varphi_\vartheta = \mathrm{G}((blue \Rightarrow \mathrm{GF}blue) \wedge (red \Rightarrow \mathrm{GF}red))$。在 G_ϑ 上的任意历史之后，图 3 中的调度器处于 q_b（分别为 q_r）当且仅当曾访问过 q_1（分别为 q_2），随后执行匹配的动作，从而确保 φ_ϑ 几乎必然成立。相比之下，任何 k-序贯策略在经过 k 步后，都必须在 q_4 执行相同的动作，无论发生的是 q_1 还是 q_2，因此 φ_ϑ 不可能几乎必然成立。完整证明见附录 A.1。□

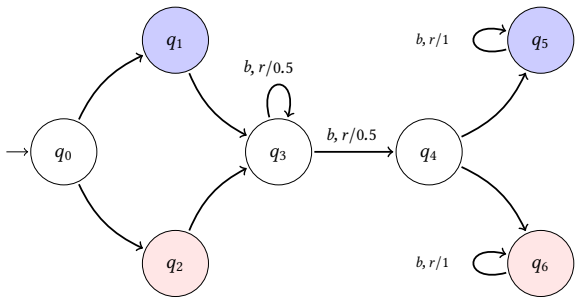

图 2：一个马尔可夫决策过程 $\mathcal{G}_\vartheta$，其中基于调度器的策略比顺序策略更一般。

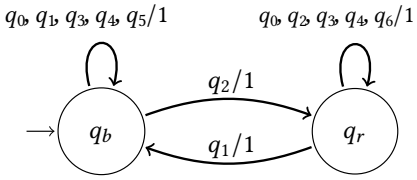

图 3：强制执行φ_g 于 $\mathcal{G}_g$上的双状态调度器。

我们强调，我们的示例使用的是确定型且完全信息的、基于调度器的策略。Cristau 等人 [20] 详细说明了随机化如何产生更一般的策略，因此我们选择将这些策略称为“一般策略”，或者直接称为“策略”。我们注意到，这两个定义在无限记忆策略上是一致的。

基于观测的策略与有界回溯策略并非互斥。我们将 $Str_{a,n}$ 视为智能体 a 的具有有界回忆 n 的基于观测的策略集合，并将 $Str_n = \cup_{a\in \mathrm{Ag}} Str_{a,n}$。

备注 1。 本文稍后用于建立复杂度结果的技术既可应用于有界回忆策略，也可应用于无记忆策略，并会得到相同的复杂度界。由于有界回忆涵盖了无记忆策略的情形且更为一般，我们将重点讨论有界回溯策略。

5 概率 ATL

接下来，我们回顾概率交替时间时序逻辑（PATL）的语法 [6], , 其定义如下：

定义 *5.1.* PATL 的语法由以下语法给出

$$\varphi ::= p \mid \varphi \vee \varphi \mid \neg\varphi \mid \langle\!\langle C\rangle\!\rangle^{\bowtie d}(\mathbf{X}\varphi) \mid \langle\!\langle C\rangle\!\rangle^{\bowtie d}(\varphi \mathbf{U} \varphi) \mid \langle\!\langle C\rangle\!\rangle^{\bowtie d}(\varphi \mathbf{R} \varphi)$$

其中 $p \in \mathrm{AP}$、 $C \subseteq \mathrm{Ag}$、 d是 $[0,1]$, 中的一个有理常数，并且 $\bowtie \in \{\leq, <, >, \geq\}$。

这些算子的直观含义如下：“下一步”**X**、“释放”**R** 和“直到” **U** 是标准的时序算子。我们使用常见的语法糖$\mathbf{F}\varphi := \top \mathbf{U} \varphi$ 和$\mathbf{G}\varphi := \bot \mathbf{R} \varphi$ 表示时序算子。$C^{\bowtie d}\varphi$ 断言存在一种基于观察的策略，使联盟 C 能够协作实现 φ，其概率与常数 d之间满足关系 $\bowtie$。

在介绍语义之前，我们先说明如何定义结果上的概率空间。

结果上的概率空间。 我们将策略组合定义为 $\boldsymbol{\sigma} = (\sigma_a)_{a\in \mathrm{Ag}}$，其从状态 s 出发的输出 是博弈路径 π ，该路径始于状态 s，并通过让每个智能体在 $\boldsymbol{\sigma}$ 中遵循其策略而延伸。形式上， $\pi_0 = s$，且对于每个 $k \geq 0$，都存在 $\boldsymbol{c}_k \in (\sigma_a(\pi_{\leq k}))_{a\in \mathrm{Ag}}$，使得 $\pi_{k+1} \in \delta(\pi_k, \boldsymbol{c}_k)$。

策略组合 $\boldsymbol{\sigma}$与状态 s所产生的（可能无限的）输出集合记为 $Output(\boldsymbol{\sigma}, s)$。

给定系统 G、策略组合 $\boldsymbol{\sigma}$ 和状态 s ，可诱导出一个无限状态马尔可夫链 $M_{\boldsymbol{\sigma}, s}$其状态为博弈路径的有限前缀（即历史），这些路径属于 $Output(\boldsymbol{\sigma}, s)$。给定 $Output(\boldsymbol{\sigma}, s)$ 中的两个历史 h和 hs'，$M_{\boldsymbol{\sigma}, s}$中的转移概率定义为 $p(h, hs') = \sum_{\Pi_{a\in \mathrm{Ag}} \mathrm{Ac}\ \sigma(\boldsymbol{c})(h) \times \delta(\mathrm{last}(h), \boldsymbol{c})(s')}$。该马尔可夫链 $M_{\boldsymbol{\sigma}, s}$在其无限路径集合 [34], 上诱导出一个规范概率空间，因此也在 $Output(\boldsymbol{\sigma}, s)$ 中诱导出规范概率空间。[1]

给定记忆界为n的联盟策略 $\boldsymbol{\sigma}_C \in \prod_{a\in C} Str_{a,n}$， 从状态 $s \in \mathrm{St}$ 出发， $\boldsymbol{\sigma}_C$的可能结果集合为$out_C(\boldsymbol{\sigma}_C, s) = \{Output((\boldsymbol{\sigma}_C, \boldsymbol{\sigma}_{-C}), s) : \boldsymbol{\sigma}_{-C} \in \prod_{a\in \mathrm{Ag}_{-C}} Str_a\}$，即玩家在 C 中遵循策略 $\boldsymbol{\sigma}_C$ 时所能确保的路径；也就是说，对于每个 $a \in \mathrm{Ag}$，玩家 a都在 $\boldsymbol{\sigma}_C$ 中遵循策略 σ_a。我们用 $\mu_s^{\boldsymbol{\sigma}_C}$表示由 $out_C(\boldsymbol{\sigma}_C, s)$诱导的测度。

定义 *5.2*。 PATL 公式在随机系统 G 和路径 π 中进行解释。令 $b \in \mathbb{N}_+ \cup \{\infty\}$。记为满足关系 $\models_b$ 的、有 b-记忆界的 PATL 语义定义如下：

$$
\begin{aligned}
&\mathcal{G}, \pi \models_b p && \text{iff } p \in v(\pi_0)\\
&\mathcal{G}, \pi \models_b \neg\varphi && \text{iff } \mathcal{G}, \pi \not\models_b \varphi\\
&\mathcal{G}, \pi \models_b \varphi_1 \vee \varphi_2 && \text{iff } \mathcal{G}, \pi \models_b \varphi_1 \text{ or } \mathcal{G}, \pi \models_b \varphi_2\\
&\mathcal{G}, \pi \models_b \langle\!\langle C\rangle\!\rangle^{\bowtie d}\varphi && \text{iff } \exists \boldsymbol{\sigma}_C \in \prod_{a\in C} Str_{a,b}\\
&&& \quad \text{s.t. } \forall \mu_{\pi_0}^{\boldsymbol{\sigma}_C} \in out_C(\boldsymbol{\sigma}_C, \pi_0),\\
&&& \quad \mu_{\pi_0}^{\boldsymbol{\sigma}_C}(\{\pi' : \mathcal{G}, \pi' \models_b \varphi\}) \bowtie d\\
&\mathcal{G}, \pi \models_b \mathbf{X}\varphi && \text{iff } \mathcal{G}, \pi_{\geq 1} \models_b \varphi\\
&\mathcal{G}, \pi \models_b \varphi_1 \mathbf{U} \varphi_2 && \text{iff } \exists k \geq 0 \text{ s.t. } \mathcal{G}, \pi_{\geq k} \models_b \varphi_2 \text{ and}\\
&&& \quad \forall j \in [0,k).\ \mathcal{G}, \pi_{\geq j} \models_b \varphi_1\\
&\mathcal{G}, \pi \models_b \varphi_1 \mathbf{R} \varphi_2 && \text{iff } \forall k \geq 0 \text{ s.t. } \mathcal{G}, \pi_{\geq k} \models_b \varphi_2 \text{ or}\\
&&& \quad \exists j \in [0,k).\ \mathcal{G}, \pi_{\geq j} \models_b \varphi_1
\end{aligned}
$$

注释 2。 在恰好包含一个智能体的系统上定义的 PATL，对应于具有有界记忆策略的部分可观测马尔可夫决策过程上的 PCTL；我们所依据的文献几乎完全聚焦于 PCTL 中仅限于可达性目标的片段。我们的目标是在多智能体系统设置下处理所有具有有界记忆的 PATL。

定义 *5.3*。 给定系统 G、状态 $s \in \mathrm{St}$、记忆界$b \in \mathbb{N}_+$ 以及 PATL 中的公式 φ，PATL 的模型检测问题 在于判定 G $s \models_b \varphi$ 是否成立。

命题 5.4（出自 [27]）。 使用无限记忆策略的PATL 模型检测问题是不可判定的。

证明。对于无限记忆，基于调度器的策略与经典序贯策略的定义相同，二者都是无界函数 $\sigma : \mathrm{obs}_a(\mathrm{Hist}) \to \mathrm{d}(\mathrm{Ac})$。因此，我们可以直接应用已知结果：在部分观测下，具有无限记忆的 PATL 是不可判定的 [27]。□

在本文余下部分，我们将假设 $b \in \mathbb{N}_+$。

6 鲁棒性

参数化系统。 在参数化系统中，转移概率由关于有限变量集 X 的方程取代。因此，我们可以考察对于 X 上所有能产生定义良好的概率集的赋值，某个性质是否成立。我们采用与 [33] 类似的定义。

[1]这是一个经典构造，参见例如 [10, 18]。

一个 参数化随机系统 （或参数系统）$\mathcal{G}$ 是一个元组(St, L $X, \delta, v,$ $(\text{obs}_a)_{a\in\text{Ag}}, \kappa$) 其中

- St、L、 v 和 $(\text{obs}_a)_{a\in\text{Ag}}$ 的定义与随机系统相同；
- X 是一个有限的实值参数集合；
- 对于每个状态 $s \in \text{St}$ 和每个动作 $\boldsymbol{c} \in \text{L}(s)$，参数化随机转移函数 δ 给出从状态 s 转移到所有 $s' \in \text{St}$ 的（条件）概率，前提是每个玩家 $a \in \text{Ag}$ 都执行动作 $\boldsymbol{c}_a$。该概率是关于 X 的有理函数，形式化表示为 $\delta(s, \boldsymbol{c})(s') \in \mathbb{Q}[X]$；并且
- κ是一个形如 $A \bowtie B$ 的多项式约束集合，其中 $A, B \in \mathbb{Q}[X]$ 以及 $\bowtie \in \{<, >, \leq, \geq, =\}$。之后，我们将利用这些约束来描述不同类型的扰动。

对于参数系统 G，赋值 $\text{Val} : X \to \mathbb{R}$ 若将每个 $x \in X$ 替换为 $\text{Val}(x)$ 后得到随机系统 G[Val]，则称其定义良好。将 Val 的陪域扩展为关于 X 的有理函数后，我们可以将定义良好的赋值形式化定义为满足以下条件：(i) 概率非负：对于每个 $s, s' \in \text{St}$ 和 $\boldsymbol{c} \in \text{L}(s)$，有 $\text{Val}(\delta(s, \boldsymbol{c}, s')) \geq 0$；(ii) 转移会产生分布：对于每个 $s, \in \text{St}$ 和 $\boldsymbol{c} \in \text{L}(s)$，有 $\Sigma_{s'\in\text{St}}\text{Val}(\delta(s, \boldsymbol{c}, s')) = 1$。

如上所述，参数化部分可观测马尔可夫决策过程或参数化部分可观测马尔可夫决策过程被定义为仅含一个智能体的参数系统，参数化马尔可夫决策过程或参数化马尔可夫决策过程被定义为具有完全信息的参数化部分可观测马尔可夫决策过程，而参数化马尔可夫链或参数化马尔可夫链则被定义为对每个 $s \in \text{St}$均满足 $|\text{L}(s)| = 1$ 的参数化马尔可夫决策过程。

例 *6.1*（*River*（续））。现假设这条河的水质不如先前假定的那样稳定。我们指出这种不稳定性中存在某种依赖关系，并用参数加以表示。部分转移概率现在包含扰动 x，且所有这些概率会一同变化。我们在例 3.1 的系统 G_{river} 中采用如下由 x 参数化的扰动族，称之为 G^p_{river}，并在图 4 中加以说明。

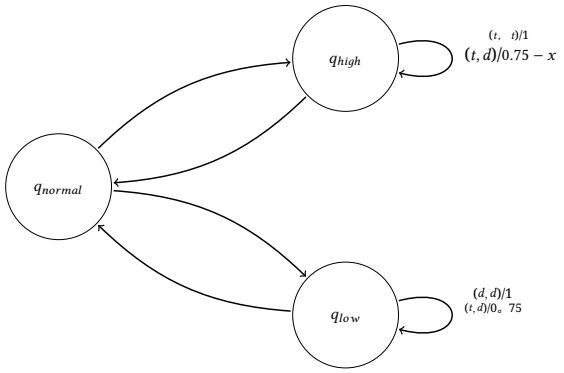

图 4：参数化系统 $\mathcal{G}^p_{river}$ 展示了这样一种情况：我们不确定某些转移，而这些转移可能会根据 $x \in [0, 0.25]$ 一同发生变化。

处于状态 q_{normal} 时，仅有一家公司处理水时水质改善的概率可能略高于计划值，因此转移到 q_{high} 的概率为 $0.5 + x$。因此，为了进行补偿，在这种情况下转移到 q_{low} 的概率为 $0.5 - x$。与此同时，仅有一家公司处理水时，保持在 q_{high} 的概率可能略低于预期，但趋势相同，因此该概率为 $0.75 - x$。不过，我们可以确认，只要 x 保持在 0 到 0.25 之间，一家公司单独处理水，就始终能确保在两个时间步内获得良好水质的概率至少为 0.375。

根据先前的工作 [15, 22, 26, 47], ，我们所考虑的鲁棒策略概念需要处理精确概率方面的不确定性。这种不确定性通过系统可能遭遇的扰动 来建模。我们的首要目标是验证某个策略是否能够实现目标，即使系统可能遭遇 $\varepsilon \in [0, 1]$ 的扰动。我们首先介绍 ε-近似系统。

给定一个随机系统 G $= (\text{St}, \text{L}\ \delta, \ell, (\text{obs}_a)_{a\in\text{Ag}})$，以及 $\varepsilon \in [0, 1]$, ，我们定义 ε-近似参数化系统 $\text{G}_\varepsilon = (\text{St}, \text{L}, X, \delta', \ell, (\text{obs}_a)_{a\in\text{Ag}}, \kappa)$。其中，对于每个 $s, s' \in \text{St}$ 和 $\boldsymbol{c} \in \text{L}(s)$，每个转移概率 $\delta'(s, \boldsymbol{c}, s') \in X$都是一个新变量。我们向$\kappa$添加约束： $\delta'(s, \boldsymbol{c}, s') - \delta(s, \boldsymbol{c}, s') \leq \varepsilon$且 $\delta(s, \boldsymbol{c}, s') - \delta' s\ \boldsymbol{c}\ s' \leq \varepsilon$ $|\delta\ s\ \boldsymbol{c}\ s' - \delta'\ s\ \boldsymbol{c}\ s'| \leq \varepsilon$ $(, ,)$，或者等价地，$(, ,) (, ,)$。因此，与 G 相比，G_ε 中的每个概率最多可能存在 ε 的误差。区间马尔可夫决策过程是 ε-近似参数化系统的一种特殊情况；更准确地说，区间马尔可夫决策过程对应于 ε-近似马尔可夫决策过程 [16]。

备注 3. 一般而言，参数可能会导致图的结构出现不连续，因为当边的概率变为 0 时，边可能会消失。虽然这可能会改变问题的复杂度，但我们用作 [16, 33] 构建模块的两个结果并非如此。因此，我们无需检查赋值参数是否会改变图的结构。

例 *6.2*（*River*（续））。现在，我们考虑例 3.1 中系统 G_{river} 的区间扰动 $\epsilon = 0.05$。对于每个可能的状态 q、 q' 以及联合动作（x， y），0.05-近似参数化系统 $\text{G}_{0.05,river}$ 允许转移概率发生变化，且变化幅度与 G_{river} 中的转移概率相差至多 0.05。因此，概率 0.5 可以替换为 $[0.45, 0.55]$ 中的任意值。图 5 展示了 $\text{G}_{0.05,river}$ 的一种可实现化结果。所得系统记为 G'_{river}，它与 G_{river} 相比有一些显著差异。首先，当两家公司都排放废水时，从 q_{high} 出发的转移不再是确定型的。该转移仍以较高概率到达状态 q_{normal}（即水质降至正常水平），但也有较小概率使水质降至低水平，从而转移至状态 q_{low}。其次，在状态 q_{normal} 下，若只有一家公司处理水质，那么水质下降（转移至状态 q_{low}）的可能性高于水质上升（转移至状态 q_{high}）的可能性。显然，与原系统 G_{river} 相比，在 G'_{river} 中从另外两个状态到达状态 q_{low} 的概率更高。请注意，并非 $\text{G}_{0.05,river}$ 的每一种可实现化结果都如此，因为它还允许对转移概率进行其他修改。

现在我们介绍 PATL 的鲁棒模型检测问题。

定义 *6.3.* 给定一个参数化系统 G、状态 $s \in \text{St}$、PATL 中的公式 φ 以及记忆界限 b，*PATL* 的参数化模型检测问题旨在判定对于每个定义良好的赋值 Val，$\text{G}[Val], s \models_b \varphi$是否成立。

当 G 是一个 ε-近似系统时，我们称之为 PATL 的 ε-鲁棒模型检测问题。

我们还考虑假设受扰动转移的数量固定不变的情况。事实上，在许多情况下，我们可以认为

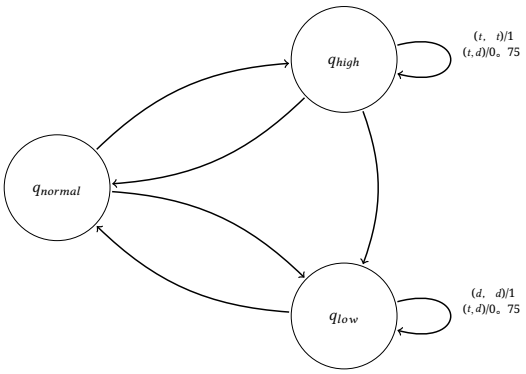

图 5：系统 $\mathcal{G}'_{river}$，由对 0。05-近似系统 $\mathcal{G}_{0_{\circ}05,\ river}$ 的可实现化得到。

只有少数关键组件的行为存在不确定性。正如我们将看到的，这一假设有助于大幅降低模型检测复杂度。

7 PATL 鲁棒模型检测

我们首先考虑规范 φ 定义可达性目标的情形。更准确地说，我们关注这样的情形：可达性规范 φ 对智能体的所有（任意）策略都成立，形式化表示为：$\varphi = \emptyset^{\bowtie d}\mathbf{F}R$，其中 d 是 $[0,1]$中的一个有理常数，$\bowtie \in \{\leq, <, >, \geq\}$，而 R 是原子命题的布尔组合。这等价于考察一个公式是否对参数化马尔可夫决策过程中的所有可能策略都成立。要获得存在可达性方面的结果，可以利用对偶性：存在一种策略以概率 $> d$ 到达目标状态 R，当且仅当我们能够证明并非所有策略都以概率 $\leq d$到达 R。有关参数化可达性的更多结果，请参见 [33]。

我们从 ε-近似马尔可夫决策过程开始。由于它们是区间马尔可夫决策过程的一种特例，我们可以使用以下结果：

定理 7.1。 *[根据 [16]]*，以下问题可在多项式时间内解决：给定一个 ε-近似 $MDP\ \mathcal{G} = (\mathrm{St}, \mathrm{L}, X, \delta, \ell, (\mathrm{obs}_a)_{a\in\mathrm{Ag}}, \kappa)$ 和可达性目标 R，确定对于所有 b-有界策略以及所有定义良好的参数 X，R 是否能以概率 $\bowtie d$，在 $\bowtie \in \{<, >, \geq, \leq\}$ 内到达。

对于更一般的参数化系统情形，我们首先考虑参数数量有上界的情况，并扩展 [33] 中关于参数化马尔可夫决策过程的结果。

定理 7.2。 以下问题属于 **NP** ∩ **co-NP**：给定一个 参数-$MDP\ \mathcal{G} = (\mathrm{St}, \mathrm{L}, X, \delta, \ell, (\mathrm{obs}_a)_{a\in\mathrm{Ag}}, C)$，其中参数数量小于一个固定的 m，以及一个可达性目标 R，判断对于所有 b-有界策略和所有定义良好的参数 X，R 是否能以概率 $\bowtie d$ 到达。

证明。我们首先证明该问题属于**co-NP**。对偶问题，即判断是否存在一个参数和一个策略，使其确实以概率 $\bowtie d$ 满足可达性概率要求，根据 [33]，的命题 15，该问题属于**NP**；因此，我们的问题属于**co-NP**。

为了证明我们的问题属于**NP类**，我们首先指出，我们试图在一个马尔可夫决策过程中寻找满足可达性性质的策略：我们将自己限制在无记忆策略范围内（根据引理 10.102（见 [5]）。我们采用与 [33]中的定理12相同的技巧：我们猜测一个代表最坏情况的策略，该策略要么最小化（对于 $\geq d$ 和 $> d$），要么最大化（对于 $\leq d$ 和 $< d$）到达 R的概率，并检查这个最坏情况策略是否仍然最优。借助贝尔曼方程，我们将这个最优性约束转化为一个含固定数量参数的 $\exists\mathbb{R}$ 公式 φ。由于我们试图判定，对于 X上所有定义良好的赋值，φ 是否成立，因此我们得到一个含固定数量参数的 $\forall\exists\mathbb{R}$ 公式，可在**P**中进行检验。□

最后，我们考虑参数数量任意的情形。从现在起，我们开始使用涉及实数理论的复杂性类。多项式层次可以理解为寻找使布尔值公式成立的取值，而实数理论研究实值公式，由此得到其自身的复杂性类。特别是，$\forall\mathbb{R}$ 包含判断一个公式对其变量的所有可能实数赋值是否都为真的问题。对于整数 k，类 $\Pi_k^{\mathbb{R}}$ 表示判断以 $\forall$ 开始且具有 $k-1$ 次量词交替的公式的问题。我们在附录中对这一理论作进一步介绍。

定理 7.3。 以下问题属于 $\forall\mathbb{R}$：给定一个参数化-马尔可夫决策过程 $\mathcal{G} = (\mathrm{St}, \mathrm{L}, X, \delta, \ell, (\text{观测}_a)_{a\in\mathrm{Ag}}, \kappa)$，以及一个可达性目标$R$，判断对于所有 b 有界策略和所有定义良好的参数 X，R 是否能以概率 $\bowtie d$ 到达。

现在我们可以取一个系统 G，并考虑在不同的鲁棒性定义下对 PATL 进行模型检测。我们先看一个例子，然后再考虑 ε-鲁棒性的情形，并从一个稍后会进一步探讨的子情形入手。

例*7.4*（*River*（续）。）回到贯穿全文的例子，令 b 表示记忆界。下面的公式表示，联盟 C 拥有一种策略，能够保证河流水质在下一步状态中为高的概率至少为 0。5：$C^{\geq 0.5}\mathbf{X}high$。

在系统 G_{river}中，从状态 q_{normal}出发，对于任意非空联盟，该公式在记忆不超过 b的条件下为真，即 G_{river}，$q_{normal} \models_b \bigwedge_{C\neq\emptyset} C^{\geq 0.5}\mathbf{X}high$。然而，在修改后的系统 G'_{river}（图 5）中情况并非如此，因为单个公司无法以足够高的概率保证下一步到达q_{high}。因此，

$$\mathcal{G}'_{river}, q_{normal} \not\models_b \bigwedge_{C\neq\emptyset} \langle\!\langle C\rangle\!\rangle^{\geq 0.5}\mathbf{X}high$$

对于 $\mathrm{G}_{0.05,river}$ 的所有可实现化，双方公司都应合作，确保河流在下一状态保持清洁的概率至少为 0.95。因此，

$$\mathcal{G}_{0.05,river}[\mathrm{Val}], q_{normal} \models_b \langle\!\langle \mathrm{Ag}\rangle\!\rangle^{\geq 0.95}\mathbf{X}high$$

对于任意赋值 Val ．

下文中，如果公式 φ 包含一个子公式 $C^{\bowtie d}\psi$，其中 ψ 本身包含一个联盟算子，则称该公式具有联盟算子嵌套。我们将这类公式用作基础情形，并以此构建适用于完整逻辑的判定过程。

引理 7.5。 以下问题属于**NP类**：给定一个 ε近似系统$\mathcal{G} = (\mathrm{St}, \mathrm{L}, X, \delta, \ell, (\mathrm{obs}_a)_{a\in\mathrm{Ag}}, \kappa)$，在使用联盟的确定型 b有界策略时，对PATL 公式 φ 进行 ε-鲁棒模型检测，且不含联盟算子嵌套。

由此我们得出关于鲁棒模型检测问题的结论：

定理 7.6。以下问题属于**NP类**$^{\mathrm{NP}} = \Sigma_2^{\mathrm{P}}$：给定一个 ε-近似系统$\mathcal{G} = (\mathrm{St}, \mathrm{L}\ X, \delta, \ell, (\mathrm{obs}_a)_{a\in\mathrm{Ag}}, \kappa)$，使用确定型 b-有界策略，对联盟的PATL 公式 φ 进行 ε-鲁棒模型检测。

对于参数数量固定的参数化系统，存在类似引理 7.5 的结果，但该问题属于**NP类**∩**co-NP类**：

引理 7.7。以下问题属于**NP类**：给定一个参数-系统$\mathcal{G} = (\mathrm{St}, \mathrm{L}\ X, \delta, \ell, (\mathrm{obs}_a)_{a\in\mathrm{Ag}}, C)$，其中参数数量小于某个固定值 m，对不含联盟算子嵌套的PATL 公式φ 进行参数化模型检测，并对联盟使用确定型 b-有界策略。

于是，我们可以得出与定理 7.6 相同的结论。

定理 7.8。以下问题属于**NP类**$^{\mathrm{NP}} = \Sigma_2^{\mathrm{P}}$：给定一个参数-系统$\mathcal{G} = (\mathrm{St}, \mathrm{L}\ X, \delta, \ell, (\mathrm{obs}_a)_{a\in\mathrm{Ag}}, C)$，其中参数数量固定且小于 m，对PATL 公式进行参数化模型检测，并对联-盟使用确定型 b-有界策略。

现在我们考虑参数任意的情况：

定理 7.9。以下问题属于 $\exists\forall\exists\mathbb{R} = \Sigma_3^{\mathbb{R}}$：给定系统 $\mathcal{G} = (\mathrm{St}, \mathrm{L}\ X, \delta, \ell, (\mathrm{obs}_a)_{a\in\mathrm{Ag}}, C)$ 以及记忆界 $b \in \mathbb{N}_+$，对一个PATL 公式 φ 进行参数化模型检测，使用随机化的 b-有界策略。

8 应用讨论

信息物理系统是一个天然的应用领域，在该领域中，部分可观测条件下的鲁棒策略推理不可或缺 [49]。例如智慧城市、交通基础设施和工业自动化，其中软件控制器与物理环境交互，而其动态特性只能近似获知，且故障或攻击可能扰乱系统的预期行为。在这类场景中，系统需要具备鲁棒性，以应对噪声和干扰，从而高效运行，并避免攻击者利用潜在风险和威胁 [53]。

一个典型场景是智慧城市中的绿色建筑。建筑之间相互通信和协调，以高效共享能源和水等资源 [48]。每栋建筑都有私有信息（例如资源水平或用户偏好）和局部目标（例如降低成本或为社区贡献资源），而整个系统则应维持全局服务质量和安全约束 [49]。这种合作行为可以用 PATL 规范来描述。例如，公式 $C_{\geq 0.8}\mathbf{G}\left(\bigwedge_{a\in C} \neg\text{能源低}_a\right)$ 表示一个由建筑组成的联盟C 拥有一种基于观测的有界记忆策略，能够确保其成员均不会达到不可接受的低能源水平，且该保证成立的概率至少为 0.8。接下来，我们的鲁棒语义考察：在规定的不确定性集合内，对转移概率的所有 实例化，该保证是否都成立。例如，这可以描述太阳能电池板效率的偏差、供暖需求预测的不准确性，或根据新数据获得的模型更新。

同一框架还支持面向安全性的信息物理基础设施规约 [53]。将攻击者与防御者之间的交互建模为随机多智能体系统，便可以表达诸如 $\neg\{a\}_{\geq 0.2}\mathbf{F}(\text{访问}_a \wedge \neg\text{安全检查}_a)$之类的性质，其表明攻击者 a 并不拥有以至少0.2的概率获得未经授权的访问的策略。随后，鲁棒模型检测验证：即使成功攻击或检测的概率发生变化，该结论仍然成立。

从建模的角度来看，我们的鲁棒随机多智能体系统可视为经典不确定模型（如区间模型和参数化马尔可夫决策过程）的多智能体推广，并加入了联盟策略算子以及基于观测的有界记忆策略。表 1 中的复杂度结果表明，在这些一般条件下，鲁棒验证仍然是可判定的，其复杂度也与现有验证技术相容。这表明，我们的框架可为未来在现实信息物理应用中开展鲁棒策略的工具支持分析提供语义基础。

9 结论

本文研究了随机多智能体系统中智能体策略的鲁棒性验证问题，其中转移概率具有不确定性。我们聚焦于由有限自动机表示的基于观测的有界记忆策略，提出了 PATL 模型检测问题的鲁棒变体，并针对三种不同的扰动概念给出了其复杂度上界：每次转移的偏差不超过 ε、参数数量固定的参数化模型，以及参数数量无界的参数化模型。在定量方面，我们的结果阐明了多智能体系统中可验证鲁棒性的计算成本。具体而言，我们证明，即使存在联盟、不完全信息和参数化不确定性，鲁棒策略推理仍然是可判定的，同时也指出参数数量无界会引发显著的复杂度跃升。从概念上看，我们的工作将此前相对独立的几条研究脉络联系起来：策略逻辑中随机多智能体系统的验证、不确定概率的区间模型与参数化模型，以及鲁棒策略综合。

我们尚未解决下界问题；尽管有些下界高得不切实际，但我们认为，为 PATL 的鲁棒模型检测求得下界是可行的。作为未来的工作，我们计划研究如何计算给定策略的鲁棒性水平，也就是该策略仍能成功时所能承受的最大扰动。这将引出一个问题：如何在随机多智能体系统中找到鲁棒性最高的策略。

致谢

本研究得到了 DFG 项目 POMPOM (KA 1462/6-1) 和 ANR-22-CE48-0012 的资助。

参考文献

[1] Mohamadreza Ahmadi、Murat Cubuktepe、Nils Jansen、Sebastian Junges、Joost- Pieter Katoen 和 Ufuk Topcu。2018。《我们为网络欺骗而进行的部分可观测博弈》。*CoRR* abs/1810.00092（2018）。[2] Rajeev Alur、Thomas A. Henzinger 和 Orna Kupferman。2002。《交替时间时序逻辑》。*ACM*期刊 49卷5期（2002），672–713。https://doi.org/10. 1145/585265.585270[3] Benjamin Aminof、Marta Kwiatkowska、Bastien Maubert、Aniello Murano 和 Sasha Rubin。2019。《概率策略逻辑》。载于*IJCAI 2019*会议论文集。32–38。

[4] Christel Baier、Christian Hensel、Lisa Hutschenreiter、Sebastian J unges、Joost- Pieter Katoen 和 Joachim Klein。2020。《参数化马尔可夫链：PCT L复杂性与无分数高斯消元》。Information and Computation 272（2020），104504。[5] Christel Baier 和 Joost-Pieter Katoen。2008。《模型检验原理》。麻省理工学院出版社。[6] Francesco Belardinelli、Wojciech Jamroga、Munyque Mittelmann 和 Aniello Murano。2023。《随机环境中遗忘型智能体的战略能力》。载于KR 2023会议论文集。726–731。[7] Francesco Belardinelli、Wojtek Jamroga、Munyque Mittelmann 和 Aniello Murano。2024。《具有遗忘策略的随机多智能体系统验证》。载于AAMAS 2024会议论文集。160–169。[8] Francesco Belardinelli、Alessio Lomuscio、Aniello Murano 和 Sasha Rubin。2020。《利用策略逻辑验证具有公开动作的多智能体系统》。Artificial Intelligence 285（2020）。[9] Michael Benedikt、Rastislav Lenhardt 和 James Worrell。2013。《区间马尔可夫链的线性时序逻辑模型检验》。载于TACAS 2013会议论文集，第7795卷。施普林格，32–46。https://doi.org/10.1007/978-3-642-36742-7_3[10] Raphaël Berthon、Nathanaël Fijalkow、Emmanuel Filiot、Shibashis Guha、Bastien Maubert、Aniello Murano、Laureline Pinault、Sophie Pinchinat、Sasha Rubin 和 Olivier Serre。2020。《具有定性语义的交替树自动机》。ACM计算逻辑汇刊22卷1期（2020），1–24。[11] Raphaël Berthon、Joost-Pieter Katoen、Munyque Mittelmann 和 Aniello Murano。2024。《随机多智能体系统中的自然战略能力》。载于AAAI 2024会议论文集。17308–17316。[12] Raphaël Berthon、Bastien Maubert、Aniello Murano、Sasha Rubin 和 Moshe Y Vardi。2021。《不完美信息策略逻辑》。ACM计算逻辑汇刊22卷1期（2021）。[13] John F. Canny。1988。《PSPACE中的一些代数与几何计算》。载于第20届ACM年度计算理论研讨会论文集。ACM，460–467。https://doi.org/ 10 1145 62212 62257 . / . Petr Cermák、Alessio Lomuscio、Fabio Mogavero 和 Aniello Murano。2018。《针对SLK规范的多智能体系统实用验证》。Inf. Comput. 261（2018），588–614。[15] Krishnendu Chatterjee、Ehsan Kafshdar Goharshady、Mehrdad Karrabi、Petr Novotný 和 Dorde Zikelic。2024。《通过随机博弈求解长期平均奖励鲁棒马尔可夫决策过程》。载于IJCAI 2024会议论文集。6707–6715。https://doi.org/ 10 24963 2024 741 . / ijcai. / Taolue Chen、Tingting Han 和 Marta Z. Kwiatkowska。2013。《区间值离散时间马尔可夫链模型检测的复杂性》。Inf. Process. Lett. 113, 7 2013 210 216 10 1016 2013 01 004 ( ), –。[17] https://doi.org/ . /J.IPL. . . Taolue Chen 和 Jian Lu。《概率交替时序逻辑与模型检验算法》。载于FSKD 2007会议论文集。35–39。[18] Edmund M. Clarke、Orna Grumberg、Daniel Kroening、Doron Peled 和 Helmut Veith。2018。《模型检验》。麻省理工学院出版社。[19] Edmund M. Clarke、Thomas A. Henzinger、Helmut Veith 和 Roderick Bloem。2018。《模型检验手册》（第1版）。施普林格出版公司。[20] Julien Cristau、Claire David 和 Florian Horn。2010。《我们如何在随机化策略中记住过去？》。载于GANDALF 2010会议论文集（EPTCS，第25卷）。30–39。[21] Murat Cubuktepe、Nils Jansen、Sebastian Junges、Joost-Pieter Katoen 和 Ufuk Topcu。2022。《用于马尔可夫决策过程参数综合的凸优化》。IEEE Trans. Autom. Control. 67卷12期（2022），6333–6348。[22] Murat Cubuktepe、Nils Jansen、Sebastian Junges、Ahmadreza Marandi、Marnix Suilen 和 Ufuk Topcu。2021。《不确定部分可观测马尔可夫决策过程的鲁棒有限状态控制器》。载于AAAI人工智能会议论文集，第35卷。11792–11800。[23] Julian D'Costa、Engel Lefaucheux、Eike Neumann、Joël Ouaknine 和 James Worrell。2021。《紧半代数集上线性动力系统逃逸问题的复杂性》。载于计算机科学数学基础国际研讨会论文集，MFCS 2021。33:1–33:21。[24] Catalin Dima 和 Ferucio Laurentiu Tiplea。2011。《不完美信息和完美回忆语义下ATL模型检测不可判定》。CoRR abs/1102.4225（2011）。[25] Michael Gene Dobbins、Linda Kleist、Tillmann Miltzow 和 Pawel Rzazewski。2023。《复杂性类 ∀∃{*R*}的完备性与面积通用性》。Discret. Comput. Geom. 70卷1期（2023），154–188。https://doi.org/10.1007/S00454- 022-00381-0[26] Ernst Moritz Hahn、Vahid Hashemi、Holger Hermanns、Morteza Lahijanian 和 Andrea Turrini。2019。《具有多目标的区间马尔可夫决策过程：从鲁棒策略到帕累托曲线》。ACM建模与计算机仿真汇刊（TOMACS）29卷4期（2019），1–31。[27] Jianye Hao、Songzheng Song、Yang Liu、Jun Sun、Lin Gui、Jin Song Dong 和 Ho-fung Leung。[n.d.]。《使用计数器抽象对分散博弈中的多智能体行为进行概率模型检验》。载于PRIMA 2012，马来西亚砂拉越州古晋（计算机科学讲义，第7455卷）。施普林格，16–30。https: //doi.org/10.1007/978-3-642-32729-2_2 [28] Wojciech Jamroga 和 Jürgen Dix. 2006. 不完全信息下的能力模型检测确实是 Delta2-完全。在第 4 届欧洲多智能体系统研讨会（EUMAS 2006）论文集。[29] Wojciech Jamroga、Damian Kurpiewski 和 Vadim Malvone. 2020. 投票协议中的自然战略能力。发表于社会技术安全与信任国际研讨会。45–62。[30] Wojciech J amroga、Vadim Malvone 和 Aniello Murano. 2019. 自然战略能力。Artificial Intelligence 277 (2019), 103170. https://doi.org/10.1016/j.artint. 2019.103170[31] Sebastian Junges. 2020. 马尔可夫模型中的参数综合。博士论文。德国亚琛工业大学。https://publications.rwth-aachen.de/ record/783179[32] Sebastian Junges、Erika Ábrahám、Christian Hensel、Nils Jansen、Joost-Pieter Katoen、Tim Quatmann 和 Matthias Volk. 2024. 马尔可夫模型的参数综合：覆盖参数空间。Formal Methods Syst. Des. 62, 1 (2024), 181–259. https://doi.org/10.1007/S10703-023-00442-X [33] Sebastian Junges、Joost-Pieter Katoen、Guillermo A. Pérez 和 Tobias Winkler. 2021. 参数化马尔可夫决策过程中的可达性复杂度。J. Comput. Syst. Sci. 119 2021 183 210 10 1016 2021 02 006 ( ), – . [34] https://doi.org/ . /J.JCSS. . . John G. Kemeny、J. Laurie Snell 和 Anthony W. Knapp. . 随机过程。载于《可数马尔可夫链》。施普林格，40–57。[35] Navdeep Kumar、Kaixin Wang、Kfir Yehuda Levy 和 Shie Mannor. 2024. s-矩形鲁棒马尔可夫决策过程的高效值迭代。发表于 ICML 2024，奥地利维也纳。https://openreview.net/forum?id=J4LTDgwAZq [36] Marta Kwiatkowska、Gethin Norman 和 David Parker. 2022. 概率模型检测与自主性。Annual Review of Control, Robotics, and Autonomous Systems 5 (2022), 385–410.[37] Marta Kwiatkowska、Gethin Norman、David Parker、Gabriel Santos 和 Rui Yan. 2022. 基于战略均衡的决策制定中的概率模型检测：进展与挑战。发表于 MFCS 2022 论文集。4–22。[38] Thomas M. Moerland、Joost Broekens、Aske Plaat 和 Catholijn M. Jonker. 2023. 基于模型的强化学习：综述。Foundations and Trends® in Machine Learning 16, 1 (2023), 1–118. [39] Aniello Murano、Daniel Neider 和 Martin Zimmermann. 2023. 鲁棒交替时间时序逻辑。发表于 JELIA 2023 论文集。施普林格，796–813。[40] Alberto Puggelli、Wenchao Li、Alberto L. Sangiovanni-Vincentelli 和 Sanjit A. Seshia. 2013. 对具有凸不确定性的马尔可夫决策过程的 PCTL 性质进行多项式时间验证。发表于 CAV 2013 论文集。527–542。[41] Alberto Puggelli、Alberto L. Sangiovanni-Vincentelli 和 Sanjit A. Seshia. 2014. 将鲁棒策略综合应用于风险限制型可再生能源定价的概率系统。发表于 EMSOFT 2014 论文集。ACM，13:1–13:10。[42] Marcus Schaefer、Jean Cardinal 和 Tillmann Miltzow. 2024. 作为复杂性类的实数的存在理论：概要。arXiv 预印本 arXiv:2407.18006 (2024)。[43] Marcus Schaefer 和 Daniel Stefankovic. 2024. 超越实数的存在理论。Theory Comput. Syst. 68, 2 (2024), 195–226. https://doi.org/10.1007/ S00224-023-10151-X[44] Uwe Schöning. 1983. NP 中的低层与高层层次结构。J. Comput. Syst. Sci. 27, 1 1983 14 28 10 1016 0022 0000 83 90027 2 ( ), – . [45] https://doi.org/ . / - ( ) - Fu Song、Yedi Zhang、Taolue Chen、Yu Tang 和 Zhiwu Xu. . 概率交替时间 μ演算。发表于 AAAI 论文集。6179–6186。[46] Larry J. Stockmeyer. 1976. 多项式时间层次。Theor. Comput. Sci. 3, 1 (1976), 1–22. https://doi.org/10.1016/0304-3975(76)90061-X [47] Marnix Suilen、Nils Jansen、Murat Cubuktepe 和 Ufuk Topcu. 2020. 通过凸优化为不确定部分可观测马尔可夫决策过程进行鲁棒策略综合。发表于 IJCAI 2020 论文集。4113–4120。[48] Wayes Tushar、Chau Yuen、Tapan K Saha、Thomas Morstyn、Archie C Chapman、M Jan E Alam、Sarmad Hanif 和 H Vincent Poor. 2021. 面向互联社区的点对点能源系统：近期进展与新兴挑战综述。Applied Energy 282 (2021), 116131.[49] Wayes Tushar、Chau Yuen、Tapan K Saha、Sohrab Nizami、Mollah Rezaul Alam、David B Smith 和 H Vincent Poor. 2023. 从博弈论视角综述信息物理系统。IEEE Access 11 (2023), 9799–9834。[50] Wolfram Wiesemann、Daniel Kuhn 和 Berç Rustem. 2013. 鲁棒马尔可夫决策过程。Math. Oper. Res. 38, 1 (2013), 153–183. https://doi.org/10. 1287/MOOR.1120.0566[51] Michael Wooldridge. 2009. 《多智能体系统导论》。John Wiley & Sons。[52] Pian Yu、Shufang Zhu、Giuseppe De Giacomo、Marta Kwiatkowska 和 Moshe Y. Vardi. 2024. LTLf 规划中的颤抖手问题。发表于 IJCAI 论文集（2024）。[53] Quanyan Zhu 和 Tamer Basar. 2015. 信息物理控制系统的鲁棒性、安全性与韧性的博弈论方法：用于最优跨层韧性控制系统的博弈中博弈原则。IEEE Control Systems Magazine 35, 1 2015 46 65 10 1109 2014 2364710 ( ), – . [54] https://doi.org/ . /MCS. . Martin Zimmermann. . 鲁棒概率时序逻辑。Inform. Process. Lett. 188 (2025), 106522.

A 附录

A.1 命题 4.2 的证明

命题 4.2。 如下结论成立：

- 对于每个 k-顺序策略，都存在一个基于调度器的有限记忆策略，使其在相同历史下执行相同的动作。
- 存在一个马尔可夫决策过程 $\mathcal{G}_\vartheta$ 和一个线性时序逻辑 公式φ_ϑ，使得某个基于调度器的有限记忆策略以φ_ϑ 在$\mathcal{G}_\vartheta$ 上满足该公式的概率为1，但无论 k如何，都不存在满足该公式的 k-顺序策略。

证明。给定一个 k-顺序策略 $\sigma\colon \mathrm{obs}_a(O^k) \to \mathrm{d}(\mathrm{Ac})$，我们构造调度器 σ' 为$(\{O \times \{\varnothing\}\}^{k-1}, act, \Delta, start)$，其中$\Delta(q,o) = \mathrm{last}^{k-2}(q) \cdot o$，其中 $\mathrm{last}^{k-2}(q)$ 是 $k-2$ 个 q 的末尾元素组成的序列； $act(q, o) = \sigma(q \cdot o)$，若有定义则取该值，否则取任意值；而 $start(o) = \{\varnothing\}^{k-1} \cdot o$。该调度器在任意初始状态 o下的输出为$\sigma(o)$；并且根据归纳法，若 σ 和 σ' 在序列 $q \in O^*$之后具有相同的输出，则它们在 $q \cdot o$上的输出也相同。

另一方面，考虑图 2 中的系统 G_ϑ。唯一的原子命题出现在状态 q_1 和 q_5 上，它们标记为 $blue$；而 q_2 和 q_6 则标记为 red。在大多数状态下，两个动作 r和 b产生相同的结果，只有在 q_4中例外： b会到达蓝色汇点，而 r会到达红色汇点。现在考虑线性时序逻辑公式 $\varphi_\vartheta = \mathrm{G}((blue \Rightarrow \mathrm{GF}blue) \wedge (red \Rightarrow \mathrm{GF}red))$。该公式表示，如果蓝色状态哪怕只被访问一次，那么蓝色状态就会被访问无穷多次；红色状态也同样如此。考虑图 3 中的调度器：在状态 q_b执行动作 b，在状态 q_r执行动作 r。在系统 G_ϑ上的任意历史之后，当且仅当蓝色状态 q_1（分别为红色状态 q_2）曾被访问过时，调度器会处于状态 q_b（分别为 q_r），并执行相应的动作，从而以概率 1 确保φ_ϑ。另一方面，对于任意记忆界限 k，对于任意 k-顺序策略，所有在 q_3中持续 k步的历史都必须在 q_4中执行相同的动作，无论访问过的是 q_1还是 q_2；这意味着φ_ϑ无法以概率 1 得到满足。

□

我们强调，我们的例子使用的是基于调度器的确定型完美信息策略。 [20] 详细说明了随机化如何导向更一般的策略，因此我们选择将策略称为“一般策略”，或简称为“策略”。我们注意到，这两个定义在无限记忆策略上是一致的。

A.2 关于可能的 Σ_2^{P} 困难性结果的备注（定理 7.6 的后续讨论）

我们证明，对系统上的 PATL 公式进行 ε-鲁棒模型检测可在 Σ_2^{P} 内完成，见定理 7.6。一种证明其具有 Σ_2^{P} 困难性的思路，是将 2QBF编码为该问题；该问题如下：判断形如 $\exists X \forall Y B(X, Y)$ 的公式是否成立，其中 B 是集合 X 和 Y 中原子的布尔组合。

当 B 是合取范式公式时，很容易将这样的公式编码为一个包含两个智能体（名为 $a_\exists$ 和 $a_\forall$）的系统：从一个与 V 的随机子句相关联的状态开始，并让智能体 $a_\exists$依次为 X 中的变量选择赋值（状态以待选变量标记，因此该智能体不能根据子句采取不同的博弈路径）。如果某个赋值满足该子句，就转到某个获胜状态 $s_\top$；否则，让智能体 $a_\forall$ 对 Y中的变量以类似方式继续操作。然后我们检查公式 $a_\exists{}^{=1}\mathrm{F}s_\top$：当且仅当公式 $\exists X \forall Y B(X, Y)$ 成立时，智能体$a_\exists$才拥有一个能够确保到达 $s_\top$ 的（无记忆）策略。

这种思路的问题在于，对于合取范式公式，2QBF 属于 NP：可以猜测 X 的一个赋值，然后在每个子句中独立地假设 Y 取最坏的可能值，而这可以在多项式时间内完成检查。

考虑任意布尔公式 B时，上述构造不再适用。可以让智能体 $a_\exists$决定一个（无记忆且基于观测的）策略，为 X 中的变量赋值，然后遍历 B：遇到 $\vee$时，由随机性选择左侧或右侧子公式；遇到 $\wedge$ 时，则由智能体 $a_\forall$ 选择下一个子公式，直到遇到一个变量。与该变量关联的智能体决定一个赋值；如果这个赋值满足该子公式，系统就转移到状态 $s_\top$。如果智能体 $a_\forall$ 拥有不完美信息，那么 $a_\exists{}^{>0}\mathrm{F}s_\top$ 成立当且仅当 $\exists X \forall Y B(X, Y)$ 成立，但实际并非如此，因为 $a_\forall$ 可以在公式的不同部分为同一变量选择不同的赋值。

按照 PATL 的定义，显然没有办法绕过这个问题。对所考虑的问题稍作修改或许能得到一个 Σ_2^{P}-困难性结果，但不容易得到 Σ_2^{P}-完备性。

如果我们不使用 PATL，而改用 ATL* 的一种变体，就可以让智能体 $a_\exists$ 展示其选定的全部赋值，然后让智能体 $a_\forall$ 同样如此，同时规定一旦选定某个赋值，就应始终选择同一赋值（$a \Rightarrow \mathrm{G}a$）。然而，这将使我们无法使用我们的模型检测技术，因为我们无法轻易地将一个 ATL* 公式转换为可达性和不变性目标。

另一种方法是让智能体 $a_\forall$ 也使用基于观测的无记忆策略，从而使赋值的选择不能依赖于公式中的位置。这将改变判定过程：我们不能再使用引理 7.5 和定理 7.1，因为它们允许任意策略。在部分可观测马尔可夫决策过程上，判定一个可达性公式是否被所有无记忆确定性策略以至少给定的概率满足，是**co-NP类**完全的（作为 [33], 中命题 15 的存在性情形的对偶，我们将其记为命题 A.1），而对马尔可夫决策过程而言它是多项式的；这反过来会大幅提高我们所提出的模型检测技术的复杂度。

命题 A.1（源自 [33]）。对于固定参数的参数化马尔可夫决策过程，$\exists$可达性$_\star^{\bowtie}$ 属于**NP类**。

A.3 多项式层次与实数理论（定理 7.2 的背景）

多项式层次。 经典复杂性类**NP类** 可以定义为这样一组判定问题：对于其中的肯定实例，可以利用多项式规模的见证在多项式时间内验证。SAT 问题是一个典型的**NP类**完全问题。**co-NP类**是其对偶复杂性类，其中的否定实例也可以用类似方式验证。利用预言机，这一概念扩展成了一个层次。给定两个复杂性类**A** 和 **B**，如果某问题在获得**B** 类的预言机后属于**A** 类，那么这类问题记作$\mathbf{A}^{\mathbf{B}}$。

在考虑多项式层次 [46], 时，**NP类** 记作Σ_1^{P}，而**co-NP类** 记作 Π_1^{P}。对于某个 k，如果一个问题在使用Σ_k^{P} 预言机（或等价地，使用 Π_k^{P} 预言机）时属于**P**，那么它就属于 $\Delta_{k+1}^{\mathrm{P}}$。类似地，复杂性类**NP类**$^{\Sigma_k^{\mathrm{P}}}$ 记作 $\Sigma_{k+1}^{\mathrm{P}}$，而复杂性类**co-NP类**$^{\Pi_k^{\mathrm{P}}}$ 记作Π_{k+1}^{P}。这些复杂性类广泛应用于模型检测，尤其是，不完美信息和不完美记忆下的ATL模型检测是 Δ_2^p -完全的 [28]；对于采用无记忆策略的NatATL，也有类似的结果 [30]。

多项式层次可以理解为寻找布尔值公式的取值，而实数理论研究的是实值公式。

实数的存在理论，$\exists\mathbb{R}$，是这样一类问题：这些问题都可以在多项式时间内归约到判定一个一阶公式 $\Phi \equiv \exists x_1 \ldots \exists x_n\, P(x_1, \ldots, x_n)$ 的真值，其中 x_i 取值于实数 $\mathbb{R}$，而 P 是由如下形式的原子谓词构成的布尔函数：$f_i(x_1, \ldots, x_n) \geq 0$ 或 $f_i(x_1, \ldots, x_n) > 0$，其中每个 f_i 都是有理数系数的多项式。已知 $\exists\mathbb{R}$ 属于**PSPACE** 且为**NP类**困难问题 [13]。许多同时涉及部分观测和概率的问题都与这一类有关，因此近年来人们对其兴趣日益浓厚。特别是，寻找一种概率策略，使其以至少给定的概率到达目标状态，并且只使用有界记忆，是一个 $\exists\mathbb{R}$-完全问题 [31]。另请参阅 [42]，其中综述了 $\exists\mathbb{R}$。量词交替更多的类似问题也形成了一个层次。对于整数 k，$\Sigma_k^{\mathbb{R}}$ 表示判定以 $\exists$ 开头且具有 $k-1$ 次量词交替的公式的问题。因此，$\Sigma_1^{\mathbb{R}} = \exists\mathbb{R}$，且 $\Sigma_2^{\mathbb{R}} = \exists\forall\mathbb{R}$。对于以 $\forall$ 量词开头的公式，类 $\Pi_k^{\mathbb{R}}$ 与之对偶。尽管有若干问题属于 $\Sigma_2^{\mathbb{R}}$ [43]，但寻找完全问题一直颇具挑战；不过，紧致逃逸问题已被证明是这类完全问题 [23]。值得注意的是，[43] 和 [25]认为，导致从层次的第一级跃升至第二级的三种主要情形是全称扩展、不精确性和鲁棒性。我们属于第三种情形，因为我们考虑的是多个具有不完美信息、需要遵循鲁棒策略的智能体。

A.4 定理 7.1 的证明

定理 7.1 的证明将用到以下结果：

命题 A.2（引自 [16]）。考虑可达性目标时，*UMC* 与 *IMDP* 语义一致。

命题 A.3（引自 [16]）。给定一个 ε-近似的马尔可夫决策过程 $\mathcal{I}$，计算使到达目标状态的概率最大（分别为最小）的参数和策略，可在多项式时间内完成。

定理 7.1。 *[由 [16]]* 以下问题可在多项式时间内解决：给定一个 ε-近似的马尔可夫决策过程 $\mathcal{G} = (\mathrm{St}, \mathrm{L}\ X, \delta, \ell, (\mathrm{obs}_a)_{a\in\mathrm{Ag}}, \kappa)$，以及一个可达性目标 R，判断对于所有 b-有界策略和所有定义良好的参数 X, R，能否以概率 $\bowtie d$ 到达 $\bowtie \in \{<, >, \geq, \leq\}$。

证明。该结果源于以下结论：[16], where ε- 近似马尔可夫决策过程与*IMDP* 语义下的区间马尔可夫链相似。此类马尔可夫决策过程在每一步中由玩家选择下一次转移的概率分布，只要该分布保持在给定界限之内。求得一个使到达目标状态的概率最大（分别为最小）的策略，可在此类系统中用多项式时间完成，正如我们将在下文的命题 A.3 中回顾的那样。因此，判断能否以概率 d 到达目标状态也是多项式时间可解的。我们可以通过将 ε-近似马尔可夫决策过程中的状态划分为两类，构造类似的情形：一类是在其中选择下一步动作的状态，由于我们可以随机化所采取的动作，因此不受限制；另一类是在其中决定已选动作下一个状态的概率分布、且该分布须满足 ε 界限的状态。 [16], 中的命题 1（我们将在下文回顾并改述为命题 A.2）表明，对于可达性性质而言，逐步选择转移概率（即 IMDP 语义）与从一开始就固定这些概率（他们称之为 UMC 语义）是等价的，这使我们能够使用命题 A.3。我们的问题是命题 A.3 的对偶，因此可以直接使用该命题：对于一个 ε-近似马尔可夫决策过程 G，所有策略在所有赋值下都以概率$\bowtie d$ 满足一个可达性目标（$\bowtie \in \{<, >, \geq, \leq\}$），当且仅当不存在某个策略和某个赋值，使得同一可达性目标以概率 $\bar{\bowtie} d$成立，其中 $\bar{<}$ 为 $\geq$，$\bar{>}$ 为 $\leq$，$\bar{\geq}$ 为 $<$，$\bar{\leq}$ 为 $>$。□

A.5 定理 7.3 的证明

定理 7.3。 以下问题属于 $\forall\mathbb{R}$：给定一个参数-马尔可夫决策过程 $\mathcal{G} = (\mathrm{St}, \mathrm{L}\ X, \delta, \ell, (\text{观测}_a)_{a\in\mathrm{Ag}}, \kappa)$以及一个可达性目标$R$，确定对于所有 b有界策略和所有定义良好的参数 X， R 是否能以概率 $\bowtie d$到达。

证明。固定并实例化参数集后，我们得到一个马尔可夫决策过程。在马尔可夫决策过程中，如果存在一个无法到达 R的策略，则可以假设该策略是无记忆的（根据 [5]中的引理 10.102）。因此，要解决此问题，只需判断对于所有无记忆策略和所有定义良好的参数， R 是否能到达。这可以用实数上的全称理论来表达。□

A.6 引理 7.5 的证明

引理 7.5。 以下问题属于NP类：给定一个 ε-近似系统$\mathcal{G} = (\mathrm{St}, \mathrm{L}\ X, \delta, \ell, (\text{观测}_a)_{a\in\mathrm{Ag}}, \kappa)$，对公式 φ 进行 ε-鲁棒模型检测。该公式为PATL 公式，不含联盟算子嵌套，且联盟使用确定型 b有界策略。

证明。我们首先猜测一个针对该联盟的确定性策略。其规模相对于联盟算子中固定的记忆界是多项式的。固定此策略后，联盟便为空：我们得到一个 ε-近似的马尔可夫决策过程 G′，其中只剩下不属于该联盟的智能体。我们使用 [2] 中命题 5.1 和 5.2 的经典方法，将 until 和 release 算子转换为可达性。要检查参数化马尔可夫决策过程 G′ 中的智能体是否无法证伪可达性性质，需要检查对于所有策略以及界为 m 的良定义 ε-扰动，该可达性目标是否成立。根据定理 7.1，这可以在多项式时间内完成，因此该问题属于 **NP类**。□

A.7 定理 7.6 的证明

定理 7.6。 以下问题属于 **NP类**$^{\mathrm{NP}} = \Sigma_2^{\mathrm{P}}$：给定一个经 ε 近似的系统 $\mathcal{G} = (\mathrm{St}, \mathrm{L}\ X, \delta, \ell, (\mathrm{obs}_a)_{a\in\mathrm{Ag}}, \kappa)$，使用联盟的确定型 b-有界策略，对公式 φ 进行 ε-鲁棒模型检测，该公式属于 PATL 。

证明。我们先说明其成员资格，给出对这类系统和公式进行鲁棒模型检测的方法。对于 的每个子公式 ，我们引入布尔变量 $a_1 \dots a_n$ $C_i \bowtie d_i \varphi_i$ φ $i \in$,, 其 $[1, n]$ φ_i 中 。在每个子公式 中，我们用相应的命题变量替换合作算子。假设我们可以使用 **NP类** 预言机，我们证明该问题属于 **NP类**：实际上，我们可以猜测每个 $a_1, \dots, a_n$ 是否成立，并借助引理 7.5，使用我们的 NP 预言机在 $C_i \bowtie d \varphi_i$ 上于多项式时间内验证该猜测是否正确。□

A.8 引理 7.7 的证明

引理 7.7. 以下问题属于 **NP类**：给定一个参数数量少于某个固定 m的参数化系统 $=(\mathrm{St}, \mathrm{L}\ X, \delta, \ell, (\mathrm{obs}_a)^{a\in\mathrm{Ag}}, C)$，对一个不含联盟算子嵌套的PATL 公式$\varphi$ 进行参数化模型检测，并对联盟使用确定的b有界策略。

证明。猜出联盟的一个确定性策略并固定该策略后，我们得到一个参数化马尔可夫决策过程。我们再次利用 [2]将直到算子和释放算子转换为可达性问题。根据定理 7.2，对于该参数化马尔可夫决策过程，检查所有策略以及固定数量的参数是否满足可达性目标，属于**NP** ∩ **co-NP**，因此该问题属于 $\mathbf{NP}^{\mathbf{NP}\cap\mathbf{co\text{-}NP}}$，即**NP** [44]。□

定理 *7.8.* 证明方法与定理 7.6 完全相同，只需使用引理 7.7，故略去证明。

A.9 定理 7.9 的证明

定理 7.9。 </style}以下问题属于： 给定系统 G $=(\mathrm{St},$ $\mathrm{L}\ X, \delta, \ell, (\mathrm{obs}_a)^{a\in\mathrm{Ag}}, C)$ 以及一个记忆界 $b \in$ $\mathbb{N}_+$，使用随机化的 b-有界策略，对 PATL 公式 φ 进行参数化模型检测。

证明。设 φ 是一个 PATL 公式，且 G $=(\mathrm{St},\ \mathrm{L}\ X,\ \delta,\ \ell,\ (\mathrm{obs}_a)^{a\in\mathrm{Ag}}$, $C)$ 是一个参数化系统。对于 n 合作子公式 $C_i \bowtie d_i \varphi_i$，我们引入布尔变量 $a_1, \dots, a_n$ 来表示 $i \in [1, n]$。在每个子公式 φ_i 中，我们用相应的命题变量替换合作算子。判定 φ 是否成立，等价于判定是否存在这些 $a_1, \dots, a_n$ 的一致赋值，以及针对每个子公式的有界记忆策略 $\sigma_1, \dots, \sigma_n$，使得对于所有定义良好的概率转移 P_T，在每个子公式 i 中，对于联盟公式中不包含的智能体的所有策略 μ_i，φ_i 成立当且仅当 a_i 为真。对于否定情形，我们将其复制为公式 φ_i'，分别关联于 σ_i', μ_i' 以及特定的 $P_{T,i}$，以证明联盟无法实现 φ_i。由于一旦固定了有界记忆策略和转移概率，我们就得到一个具有阈值目标的经典 MDP，因此可以根据 [5]的引理 10.102，自由地假设策略 μ_i 是确定型且无记忆的，从而可以用数量关于 G 为多项式的变量来表示。形式化地，我们考虑公式

$$\exists a_1 \qquad a_i \Rightarrow, \dots, \exists \sigma_i . \forall P_T . \forall \mu_i . \varphi_i \ \wedge\ \neg a_i \Rightarrow \forall \sigma_i' . \exists P_{T\,i}' . \exists \mu_i' . \varphi_i' \qquad \varphi_i$$

$a_n \cdot \bigwedge_{i\in[1,n]}() (,)$，其中每个带有 $i \in [1, n]$ 的公式都表示：在系统 G 上遵循策略 σ_i 和 μ_i，且转移概率为 $P_{T,i}$ 并处于 ε- 鲁棒性范围内时，每个 φ_i 的赋值都是 a_i。由于没有任何量化变量是共享的，我们可以将其写成以下 $\Sigma_3^{\mathbb{R}}$ 中的公式：

$$\begin{aligned} &\exists a_1, \dots, a_n . \exists \sigma_1, \dots, \sigma_n . \forall P_T . \forall \mu_1, \dots, \mu_n . \\ &\quad \forall \sigma_1', \dots, \sigma_n' . \exists P_{T,1}', \dots, P_{T,n}' . \exists \mu_1', \dots \mu_n' . \\ &\qquad \bigwedge_{i\in[1,n]} (a_i \Rightarrow \varphi_i) \wedge (\neg a_i \Rightarrow \varphi_i') \end{aligned}$$

□