区块链系统的平均场分析

Yanni Georghiades，美国德克萨斯大学奥斯汀分校TAKASHI TANAKA，美国普渡大学Sriram Vishwanath，美国德克萨斯大学奥斯汀分校

我们提出了一种新颖的框架，通过将区块链增长建模为部分可观测随机博弈（POSG），并利用平均场近似将其简化为一组部分可观测马尔可夫决策过程（POMDP），以分析区块链共识机制。该方法将工作量证明（PoW）系统中矿工的决策过程形式化，并能够以严谨的方式考察区块选择策略，以及对由此产生的马尔可夫链进行稳态分析。通过采用平均场博弈建模，我们能够高效地刻画异步区块链网络中出现的信息不对称现象。

我们的首个主要结果是精确刻画网络延迟与PoW效率之间的权衡——PoW效率即最终进入最长链的区块比例。我们证明，在稳态下，模型中观察到的权衡与理论研究结果高度吻合，从而验证了采用平均场近似的合理性。

我们的第二项主要成果是对最长链规则（LCR）进行了严格的均衡分析。我们证明，在某些温和的假设下，LCR 是一种平均场均衡，并且在最大化 PoW 效率方面具有唯一最优性。这一成果首次为在去中心化共识协议中持续采用 LCR 提供了形式化依据，同时带来了理论验证和实践启示。

除了这些核心成果之外，我们的框架还支持对替代区块选择策略、系统动态和奖励结构进行灵活实验。它为成本高昂的测试网部署或临时分析提供了一种系统且可扩展的替代方案。虽然我们主要关注中本聪式区块链，但通过修改底层马尔可夫决策过程，该模型也足够通用，能够适用于其他架构。

这些贡献共同为研究和设计区块链共识机制奠定了严谨的基础。我们的工作提供了一种结构化、适应性强且具有坚实理论基础的视角，可用于大规模分析去中心化系统。

CCS 概念：• **应用计算 → 经济学**；• **安全与隐私 → 安全与隐私经济学**。

其他关键词和短语：工作量证明、挖矿、加密货币、最长链规则、理性、安全性、激励设计、马尔可夫决策过程

arXiv:2601.05417v1 [cs.GT] 8 Jan 2026

1 引言

区块链通过提供无需依赖可信中介的共识机制，从根本上改变了去中心化系统的设计。在这些机制中，由比特币推广的最长链规则（LCR）已成为工作量证明（PoW）系统事实上的区块选择策略。尽管 LCR 在实践中取得了成功并得到广泛采用，但其最优性的严格理论依据仍未明晰。同时，了解网络延迟和区块生成率如何影响 PoW 网络中的共识稳定性，仍是一项重大挑战，尤其是在对可扩展、高效的区块链架构需求日益增长的背景下。

在这项工作中，我们提供了一个形式化的计算框架，以解决这些根本性缺口。我们将区块链增长建模为一个部分可观测随机博弈（POSG），并通过采用平均场近似，将其化简为一组部分可观测马尔可夫决策过程（POMDP）。我们提出了一种可处理的求解方法：在完全可观测价值近似启发式方法下，迭代求解连续的POMDP。

作者地址：Yanni Georghiades，yanni.georghiades@utexas.edu，德克萨斯大学奥斯汀分校，美国得克萨斯州奥斯汀；Takashi Tanaka，tanaka16@purdue.edu，普渡大学，美国印第安纳州西拉斐特；Sriram Vishwanath，sriram@utexas.edu，德克萨斯大学奥斯汀分校，美国得克萨斯州奥斯汀。

我们的方法捕捉了去中心化网络中各代理之间区块生成、传播延迟和信息不对称的动态，并可针对任意一组系统参数或均衡策略，对由此产生的马尔可夫链进行稳态分析。我们的贡献如下：

基于部分可观测马尔可夫决策过程的区块链增长模型：我们提出了一种新颖的框架，将区块链增长建模为一组部分可观测马尔可夫决策过程，为分析代理仅能获得底层系统状态的部分观测时的区块生成和选择动态，提供了一种结构化且形式化的方法。

灵活的区块选择策略框架：我们的模型便于系统地测试和比较各种区块选择策略，克服了人工分析和成本高昂的测试网络实现所带来的局限。它还支持对区块传播和生成动态、奖励结构以及区块链架构进行修改。

延迟与PoW效率的权衡特征：通过稳态分析，我们根据模型的均衡行为，精确刻画了网络延迟与PoW效率之间的权衡，其中PoW效率指最终成为最长链一部分的区块比例。我们的分析表明，这种权衡在定量上与既有的区块链吞吐量分析模型相符 [8, 21]。这一结果证实，平均场近似可用于预测大规模PoW系统中涌现的行为，并验证了我们的框架是研究区块链共识的可靠工具。

LCR 的最优性： 我们证明，在第 3 节概述的某些温和假设下，LCR 是一种平均场均衡策略，并且在最大化 PoW 效率方面是唯一最优的。据我们所知，这是**在定义明确的策略类下，对 LCR 在去中心化共识中的最优性所作的首次正式证明**，为其在分布式共识系统中的持续应用提供了理论基础。

超越中本聪式区块链的适应性： 尽管本文主要关注中本聪式区块链，但通过修改底层状态空间和转移函数，该框架也可适用于其他架构，因此能够成为更广泛区块链研究的多用途工具。

总之，这些贡献为理解区块链性能与策略提供了全新的视角。我们的框架不仅弥合了经验启发式方法与理论分析之间的差距，还为研究中本聪式区块链之外更广泛的协议类提供了可扩展且具有普适性的基础。

2 相关工作

本节回顾三个相关领域的重要成果：比特币及其他区块链实现、单代理与多代理动态博弈，以及平均场博弈。这些领域在区块链系统的分析与建模中相互交汇，并且在我们提出的仿真框架的构建过程中各自发挥着关键作用。

2.1 比特币及其他区块链实现

比特币由中本聪于2008年推出，是首个利用区块链技术实现安全点对点交易、无需依赖中心化中介的去中心化加密货币 [19]。比特币采用的共识机制称为工作量证明（PoW），矿工竞相解决计算密集型谜题。一旦找到有效解，矿工便会将一个交易区块追加到区块链上，从而延长链。LCR 是比特币中使用的一种启发式方法，它确保网络中的节点尝试

追加到累计工作量最大的链上。然而，尽管 LCR 已展现出经验上的成功，其理论基础尚未得到充分理解。目前没有正式证明表明它是最优的，甚至无法证明在所有条件下它是可用的最佳链选择策略。

除了比特币之外，还涌现出许多其他区块链系统，例如以太坊 [3],莱特币 [23], 和 Zcash [11], ，它们各自采用不同的共识机制，并在设计上有所取舍。例如，以太坊正在从 PoW 系统转向权益证明（PoS）机制。在这种机制下，矿工由验证者取代；验证者根据其持有并愿意“质押”的加密货币数量被选出，负责提出并验证新区块。PoS 消除了 PoW 中耗能巨大的挖矿过程，同时也带来了链选择和共识策略方面的新挑战。其他区块链系统则探索了替代性架构，例如 IOTA 使用的有向无环图（DAG），这种架构摆脱了传统的线性区块链结构。

尽管区块链实现多种多样，但一个共同的挑战依然存在：如何评估不同区块链架构和区块选择策略的安全性、效率与最优性。现有研究通常采用两种方法之一：i）依靠对现实世界系统或模拟器进行实证测量，从而观察（但不进行形式化分析）各种区块链属性 [5, 18]; 或 ii）采用临时构建的理论模型，这些模型针对特定区块链架构和有限的一组区块链属性 [6–9, 21]。我们提出一个基于部分可观测马尔可夫决策过程的严谨且可推广的框架，用于研究区块链系统的平稳行为。与基于模拟的方法不同，我们的方法能够捕捉异步网络中出现的复杂信息不对称和边缘情况，并可用于确定任意一组系统参数下的均衡策略。我们的框架还可扩展，以支持对多种区块链属性和架构进行形式化分析。

2.2 马尔可夫决策过程（MDPs）

马尔可夫决策过程（MDPs）是决策理论和强化学习中成熟的框架，用于对随机环境中的决策建模 [22]。一个 MDP 是一个元组 (S，A T， R)，其中 S 是有限状态空间；A 是有限动作集； $T(S_t,\ a,\ S_{t+1})$ 是在动作 a 下从状态 S_t 转移到状态 S_{t+1} 的概率；而 $R(S_t,\ a,\ S_{t+1})$ 是在动作 a 下从状态 S_t 转移到状态 S_{t+1} 时给予的奖励。

MDP 为分析序贯决策策略提供了一种结构化方法，尤其适用于动作结果不确定且未来状态仅取决于当前状态和所选动作的问题。向区块链添加新区块的过程可以视为一个序贯决策问题，矿工必须选择要追加的最佳区块。每次选择都会影响链的未来状态并影响获得的奖励，因此 MDP 非常适合用于分析推动区块链增长的激励和策略。

现有研究已使用马尔可夫决策过程研究挖矿行为和安全漏洞，例如自私挖矿和扣留区块攻击 [12, 20]。然而，这些研究大多聚焦于特定攻击途径或区块链设计的狭窄方面。我们的工作通过提供一个用于区块链增长的通用框架，扩展了这一研究方法，使人们不仅能够评估攻击途径，还能评估链选择规则、区块传播动力学和奖励结构，并对其进行稳态分析。

2.3 随机博弈

随机博弈将马尔可夫决策过程（MDP）推广到了多智能体环境，在这种环境中，状态转移和奖励取决于所有玩家的联合动作 [24]。与 MDP 不同，

在 MDP 中，单个代理针对固定环境优化其策略；而随机博弈则刻画多个决策者之间的策略互动，每个决策者都力求最大化自身奖励。具体而言，随机博弈（S、N、A T， R）与 MDP 的不同之处在于，它考虑一组 N 代理，并将所有代理的动作和奖励作为整体进行考量。

在区块链的语境下，随机博弈已被用于证明 LCR 是挖矿博弈的纳什均衡 [13]。然而，该模型的局限在于只考虑了两名矿工，且未考虑网络传播延迟的影响。我们的模型考虑了任意数量的代理，并刻画了传播延迟与每个代理的部分观测之间的关系。

2.4 部分可观测随机博弈与马尔可夫决策过程

部分可观测随机博弈（POSG） [2] 和部分可观测马尔可夫决策过程（POMDP） [15] 通过考虑状态观测中的不确定性，扩展了随机博弈和马尔可夫决策过程。在标准随机博弈或马尔可夫决策过程中，代理能够完全了解系统状态。然而，在许多现实场景中，包括区块链系统，环境的真实状态并非完全可观测。除上述定义的随机博弈和马尔可夫决策过程的标准要素外，POSG 和 POMDP 还为每个代理定义了一个有限的观测集合 Ω^i 以及一个函数 $O^i(S_t,\ a,\ \Omega_t^i)$，用于定义每种观测的概率。也就是说，代理只能获得系统状态的部分或含噪观测，这使决策变得复杂得多。为应对不确定性，每个代理都会维护一个信念状态——即对可能状态的概率分布，并随着新观测的到来动态更新。

部分可观测马尔可夫决策过程已广泛应用于机器人学、金融、医疗保健和人工智能等领域，在这些领域中，不确定性下的决策至关重要。然而，其在区块链系统中的应用仍未得到充分探索。在区块链环境下，矿工常因网络延迟或对抗行为而只能获取部分信息。例如，矿工可能无法即时获知完整的交易池或当前的最长链，因此需要根据不完整的信息做出区块选择决策。

2.5 平均场博弈

平均场博弈（MFGs） [4, 16] 用于对大量相互作用的代理构成的场景进行建模，并提供了一个可扩展的框架，用于分析多智能体系统中的策略性决策。随着玩家数量增加，由于联合动作空间和状态转移的指数级增长，多智能体博弈往往会变得难以计算。平均场博弈通过引入平均场近似来规避这一问题：每个代理不再与单个对手互动，而是与整个群体产生的总体效应互动。

近年来，MFG 已被用于对加密货币矿工的财富分布进行建模，以研究财富异质性随时间的变化 [17] 。在我们的框架中，我们使用 MFG 对异步通信网络所导致的区块图中的信息不对称进行建模。

3 系统模型

我们考虑一个包含 N 个理性代理的离散时间区块链系统，其中每个代理独立尝试生成区块并将其添加到区块链中。为模拟现实世界区块链网络固有的信息不对称，代理仅依据各自的局部区块图做出决策；局部区块图是全局区块图的子图，只包含截至当时它们已收到的区块。

3.1 模型中的时间单位

我们的系统采用两种不同的时间单位，从而构建出一个灵活且计算上易于处理的框架，用于模拟区块生成和区块传播。

3.1.1 时间步。模型中最小的时间单位，具有固定的持续时间。时间步推动区块生成和区块传播，我们将二者建模为伯努利过程。在每个时间步中：i) 以一定概率生成一个区块；ii) 每个代理以一定概率独立接收来自全局区块图的任意区块。

然而，根据系统参数的不同，生成一个区块所需的时间步数可能很多[1]。鉴于这种潜在的差异，在时间步层面建模状态转移在计算上不可行。因此，我们引入第二种时间度量。

3.1.2 区块步。一个区块步表示生成一个区块所需的若干时间步。由于区块生成具有随机性，一个区块步包含的时间步数是随机的。状态转移以区块步为粒度发生，这意味着每个代理在每个区块步中恰好选择一个动作。区块步通常称为 t 或时间，因为区块步是博弈的时间单位。

3.2 区块生成

区块生成遵循伯努利过程。在每个时间步中，以概率α生成一个区块，其中α是可调节的系统参数，用于反映不同的区块传播条件。在给定的任一时间步中，每个代理生成区块的概率相等；生成区块后，该代理会立即将其广播给所有其他代理。每个时间步至多生成一个区块[2]。通过以这种方式构建区块生成与传播过程，我们的模型为分析各种条件下的区块链动态提供了一个简单而稳健的基础。

区块生成由以下方程描述。如果 k_α 是表示一个区块步所包含时间步数的随机变量，那么一个区块步包含 k 个时间步的概率为

$$P(k_\alpha = k) = \begin{cases} (1-\alpha)^{k-1}\alpha & \text{if } k \geq 1 \\ 0 & \text{else} \end{cases}. \tag{1}$$

更一般地，如果 k_α^y 是 y 个区块步所包含的时间步数，那么 y 个区块步包含 k 个时间步的概率为

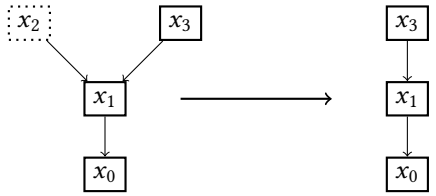
$$P(k_\alpha^y = k) = \begin{cases} \binom{k-1}{k-y}(1-\alpha)^{k-y}\alpha^y & \text{if } k \geq y \\ 0 & \text{else} \end{cases}. \tag{2}$$

3.3 区块传播模型

我们将区块传播建模为一组独立同分布（*i.i.d.*）的随机过程，其中每个过程对应一个唯一的（区块，代理）对。在每个时间步，对于代理尚未接收的任何区块 x，该代理以概率 δ 接收 x。该模型假设通信图是完全连接的，这意味着任何代理都可以从其他任意代理接收区块。

代理收到区块的概率取决于区块生成后经过的区块步数。令$H_{\alpha,\delta}$表示代理收到区块所需的区块步数

[1]例如，一个时间步可能对应工作量证明函数的一次执行。[2]可以不失一般性地作出这一假设，因为时间步可以任意缩短，从而确保在单个时间步内生成多个区块的概率可以忽略不计。



全局区块图　　局部区块图

图1。局部区块图如何从局部区块图推导而来的示例。在此示例中，代理已收到区块 x_0、x_1 和 x_3，因此其局部区块图是包含这些区块的连通子图。

那么，代理在 h 个区块步内收到区块的概率为

$$P(H_{\alpha,\delta} \le h) = \sum_{k=1}^{\infty}(1-(1-\delta)^k)P(k_\alpha^h = k). \tag{3}$$

代理恰好在区块步 h 收到该区块的概率为

$$P(H_{\alpha,\delta} = h) = P(H_{\alpha,\delta} \le h) - P(H_{\alpha,\delta} \le h-1). \tag{4}$$

由于支配区块传播的过程是无记忆的，对于代理尚未收到该区块的任一区块步，概率 $P(H_{\alpha,\delta} = h)$ 都保持不变。

4 区块链增长均衡动力学的平均场博弈

一般而言，区块链增长可以建模为 POSG

$$\mathcal{M}_{\mathsf{POSG}} = (\mathcal{S}, \mathcal{N}, \mathcal{A}, T, R, \Omega, \mathcal{O}),$$

其中，状态 $S_t \in \mathrm{S}$ 表示为以下元组

$$S_t = \left(G_t, \{B_t^i\}_{i=1,2,\ldots,N}, \{O_t^i\}_{i=1,2,\ldots,N}\right).$$

这里，G_t 是全局区块图，B_t^i 是一个 $\{0,1\}$ 向量，用于指示代理 i 已接收来自 G_t 的哪些区块；O_t^i 是一个 $\{0,1\}$ 向量，用于指示来自 G_t 的哪些区块由代理 i 生成。代理 i 并非观察 S_t，而是仅观察 Ω_t^i，它对应于截至时间 t 已接收的区块集合。这种观察称为其局部区块图，因为它对应于该代理在其局部视图中观察到的区块图。图1展示了一个局部区块图示例。

在每个区块步中，每个代理采取的动作都是从其局部区块图中选择一个想要追加到其上的区块。因此，我们将把局部区块图映射到动作的策略称为局部策略 π。稍后，我们将使用完整策略 π，它像完全可观测随机博弈中的策略一样，将完整状态映射到动作。

该模型与实际中的区块链增长高度吻合，因为它能够完整捕捉任意时刻的系统状态。然而，由于 $\mathrm{M}_{\mathsf{POSG}}$ 的状态空间基数会随着代理数量呈指数级增长，因此很难直接分析此类博弈。为了以计算上可行的方式分析区块链增长，我们采用平均场近似。

4.1 平均场近似

Mo将区块链增长建模为平均场博弈需要以下假设 s.

假设 1. 代理数量很大。

假设 2. 每个代理对区块链增长的个体影响都可以忽略不计。

假设 3. 代理身份不可区分，并且在给定的区块步骤中，所有代理生成区块的概率相等。

假设 1 在实践中相当合理，因为人们普遍认为，矿工数量越多，加密货币的去中心化程度就越高。在实践中，假设 2 和 3 并非普遍适用的运行条件，因为有些加密货币存在个体矿工或矿池控制着相当大比例计算能力的情况；但我们将在第 8 节讨论如何放宽这些假设，以便研究此类系统。不过，为简化起见，本文仅考虑计算能力的齐次分布。

在平均场近似下，区块链增长可简化为单参与者最优控制问题

$$\mathcal{M}_{\text{POMDP}} = (\mathcal{S}, \mathcal{A}, T, R, \Omega, O),$$

这与 M_{POSG} 类似，不同之处在于，状态、动作、奖励和观测仅对应单个代表性智能体（需要使用索引时，我们将不加区分地称其为 RA、代表性智能体或代理 j），而不是一组 N 代理。在此模型中，其余非代表性智能体（NRA）的观测由平均场变量 μ 以分布形式刻画，因此可以高效地表示它们，而不影响状态空间的规模。假设 NRA 遵循固定策略 π，那么，由此得到的部分可观测马尔可夫决策过程的解，可以解释为：当 NRA 遵循局部策略 π，且其局部区块图按照 μ 分布时，RA 的最佳响应。

此处，为了在后文更清晰地描述我们的模型和求解方法，我们不再采用标准记法。首先请注意，由于 S_t 包含全局区块图以及 RA 已接收哪些区块的表示，因此 Ω_t 可直接作为 S_t 的函数计算。其次，由于求解部分可观测马尔可夫决策过程本身就具有计算挑战性，我们采用了广为人知的全可观测价值近似 启发式方法（术语沿用自 [14]）。在该方法中，难以计算的部分可观测马尔可夫决策过程价值函数，使用（全可观测）马尔可夫决策过程的价值函数进行近似；后者可高效求得 [10]。我们将在第 5 节更详细地介绍这一启发式方法。鉴于此，变量 Ω 和 O 是冗余的，可以从记法中安全地省略。此外，由于我们只需求解 RA 的最佳响应，因此可将 NRA 视为系统动态的一部分，其中每一对 (π, μ) 对应一组不同的动态，并对转移函数和奖励函数产生影响。因此，我们采用以下自定义记法

$$\mathcal{M}_{\overline{\pi}} = (\mathcal{S}, \mathcal{A}, T_{\overline{\pi}}, R_{\overline{\pi}}, \mu)$$

用于指代与NRA策略 π 和均值场分布 μ相对应的部分可观测马尔可夫决策过程。我们将多智能体博弈表示为 M，它对应于所有单智能体部分可观测马尔可夫决策过程 M_{π} 的集合。为了确定 M 的均衡策略，我们迭代求解连续POMDP，直至达到不动点。

本节余下部分将详细介绍 M 的各个组成部分。 $\overline{\pi}$.

4.2 状态

时刻 t 的状态形式如下

$$S_t = (G_t, B_t, O_t),$$

其中， G_t 为全局区块图， B_t 是一个{0,1}向量，用于指示 RA 已收到 G_t 中的哪些区块； O_t 是一个{0,1}向量，用于指示 RA 生成了 G_t 中的哪些区块（这对于计算奖励函数是必要的）。为简化表示，我们对向量索引符号作如下重载：若区块 x 的索引为 k，则 $B_t[x]$ 和 $B_t[k]$ 可互换使用，表示 B_t 在索引 k 处的值。对于区块 $x \in G_t$，当且仅当代表性智能体生成了 x 时，$O_t[x] = 1$；同样，当且仅当代理截至时刻 t 已收到 x 时， $B_t[x] = 1$。

由于 M 是部分可观测的，因此不能假定代理在任何给定时刻都完全了解状态。相反，代理必须根据其局部区块图做出决策，该局部区块图仅包含代理在当前状态下从全局区块图中接收到的区块。代理 i 的**局部区块图**记为 L_t^i，定义如下：如果代理 i 尚未接收到 G_t 的根顶点，则 L_t^i 为空图；否则， L_t^i 是 G_t 中包含其根区块的最大连通子图。图1展示了如何将一个局部区块图转换为局部区块图的示例。

L_t^j 是时刻 t 代表性智能体的局部区块图，可以直接根据状态计算得出，因为 B_t 精确指定了代表性智能体从 G_t 接收到的区块。然而，代理 $i \neq j$ 的局部区块图只能通过分布来了解，其中

$$P(L_t^i = l) = \mu(l, G_t).$$

4.3 动作

在时刻 t，代表性智能体选择动作 $a_t \in \mathrm{A}$，其中动作空间是其局部区块图中的区块集合。当任一代理生成区块时，该区块会被追加到与其动作对应的区块 G_t 上。代理可用的动作仅限于其局部区块图中的区块；如果其局部区块图为空图，则没有可用动作[3]。

假设代理对状态的了解不完整，因此我们主要关注每个代理的局部策略

$$\overline{\pi} : \mathcal{G} \to \mathcal{A},$$

其中，$\overline{\pi}_i(L^i)$ 是代理 i 采取的动作，且仅取决于其对局部区块图的了解。不过，在第 5.5 节中，我们也会使用完整策略，它由决策规则给出

$$\pi : \mathcal{S} \to \mathcal{A},$$

其中， $\pi_i(S_t)$ 是代理 i 在完全了解 S_t 的情况下会采取的动作。

4.4 转移概率

每个区块步恰好发生一次状态转移，这意味着每次状态转移都会向全局区块图中追加一个区块。在给定局部策略 π 的情况下，当代表性智能体执行动作 a_t 时，从状态 S_t 转移到 S_{t+1} 的概率可分解为以下分量概率

$$T_{\overline{\pi}}(S_t, a_t, \mu, S_{t+1}) = P(G_{t+1} \mid S_t, a_t, \mu, \overline{\pi})\, P(O_{t+1} \mid S_t, a_t, \mu, \overline{\pi}, G_{t+1})\, P(B_{t+1} \mid S_t, a_t, \mu, \overline{\pi}, G_{t+1}, O_{t+1}),$$

其中， $P(G_{t+1} \mid S_t, a_t, \mu, \pi)$、 $P(O_{t+1} \mid S_t, \overline{a}_t, \mu, \pi, G_{t+1})$ 和 $P(B_{t+1} \mid S_t, a_t, \mu, \pi, G_{t+1}, O_{t+1})$ 分别表示转移至全局区块图 G_{t+1}、所有权向量 O_{t+1} 和已接收区块向量 B_{t+1} 的概率。在第 5 节中，我们将介绍一种方法，用于分别计算中本聪式区块链中每个分量的概率。

[3]这一建模决策模拟了这样一种情形：矿工或验证者与网络的步调相差甚远，以至于其决策对区块链演化的影响可以忽略不计。

4.5 奖励

尽管存在各种可能的奖励模型，我们仍定义一个通用奖励函数，奖励生成区块的代理。在中本聪式加密货币中，通常会对沿某条关键路径生成区块的行为发放奖励（在比特币中，关键路径是全局区块图中累计工作量最大的区块链）。通常，矿工需满足某些条件后才能获得奖励，例如后续区块确认数量达到足够多。为遵循这些一般约束，我们定义代表性智能体的奖励为

$$R_{\overline{\pi}}(S_t, a, S_{t+1}, \mu) = rQ(S_t, a, S_{t+1}, \mu, \overline{\pi}),$$

其中，$Q(S_t,\ a,\ S_{t+1},\ \mu,\ \pi)$ 是代理应获得奖励的关键路径区块数量，而 r 是每个区块对应的奖励。Q 的定义方式多种多样，我们将在第 5 节中结合中本聪式区块链定义 Q。

4.6 均衡概念

我们希望采用的解概念是平均场均衡 [1]。也就是说，策略 π^* 是在每个代理针对其他代理的长期平均行为优化其动作时，M 的纳什均衡。具体而言，如果 μ_{π^*} 是 M_{π^*} 在稳态下的均值场分布，那么平均场均衡策略由以下贝尔曼方程刻画

$$\overline{\pi}_j^*(L_t^j) = \arg\max_a \sum_s T_{\overline{\pi}^*}(S_t, a, s, \mu_{\overline{\pi}^*})[R_{\overline{\pi}^*}(S_t, a, s, \mu_{\overline{\pi}^*}) + \gamma V^*(s)],$$

其中，$\gamma \in [0, 1]$ 是折扣因子，且

$$V^*(s) = \max_a \sum_{s'} T_{\overline{\pi}^*}(s, a, s', \mu_{\overline{\pi}^*})[R_{\overline{\pi}^*}(s, a, s', \mu_{\overline{\pi}^*}) + \gamma V^*(s')]$$

是状态 s的预期折扣未来奖励。根据此记法，均衡的特征是.....我们求解 π^* 的方法遵循迭代式最佳响应学习过程 [4, 16]:

（1）首先，假设 NRA 遵循预先定义的策略，确定代表性智能体的最优策略。（2）接下来，我们更新 NRA 的策略，使其与代表性智能体学到的最优策略一致。（3）重复这一过程，迭代计算代表性智能体对更新后的 NRA 策略的最佳响应，直到策略收敛。

收敛表明已达到均衡策略，即没有代理有动机单方面偏离。这种迭代方法使我们能够在大规模区块链环境中近似求解最优决策问题，而在这些环境中，直接进行解析求解并不可行。在第 5.5 节中，我们将详细说明如何确定 μ_{π^*}。

5 中本聪风格区块链的均衡策略

本节将定义与中本聪式区块链相对应的状态空间、转移概率函数和奖励函数。随后，我们将定义一种定制算法，用于在任意系统参数下求解由此形成的博弈的平均场均衡。由于区块图及其关联向量具有组合性质，状态空间的基数很大，因此我们采用多种技术来限制状态空间的大小。我们给出了确定转移概率的方程，包括一种用于计算均值场分布的迭代算法，并定义了一个具体的奖励函数。

5.1 全局区块图

状态的第一个组成部分是全局区块图 G_t，它是一个有向图，其中每个顶点都是一个区块，每条边都表示对先前区块的引用。我们假设区块在功能上完全相同，也就是说，生成所有区块的难度相同，且奖励相同，但区块之间仍可相互区分（例如，根据其哈希值）。虽然单个代理的局部区块图不一定包含 G_t 中的每个区块，但代理的局部区块图中不可能有不属于 G_t 的区块。回想一下，代表性智能体的局部区块图 L_t^j 可以根据 S_t 计算得出，而其余代理的局部区块图则由 μ 在分布意义上刻画，无法直接计算。

5.1.1 图同构。 在我们考虑的中本聪式区块链中，任意图 $G \in \mathbf{G}$ 的根顶点出度为 0，其他每个顶点的出度均为 1。也就是说，除根区块外，每个区块都恰好有一条指向某个先前区块的有向边，不过多个区块可能引用同一个先前区块。我们将具有这种结构的有向图称为 中本聪图[4]。中本聪图有许多共享相同拓扑结构的分组；在假设所有区块在功能上完全相同的情况下，任何两个具有相同拓扑结构的图都会引发相同的代理响应（不考虑区块所有权）。因此，我们可以将 $\mathbf{G}$ 限定为只包含中本聪图的同构类，也就是说，只需考虑每种拓扑结构对应的一个图，而不必考虑区块和边的所有可能组合。

例如，图2展示了顶点数不超过 4 的中本聪图的所有同构类。一般而言，具有 k 个区块的中本聪图的同构类数量可以通过动态规划计算。图3说明了仅对同构类建模的好处：这实际上消除了唯一图数量中的一个阶乘项（因为在最坏情况下，每种拓扑结构内的区块都可以任意调换顺序）。

5.1.2 维护有限区块图。由于 $\mathbf{M}$ 是无限时域博弈，因此必须存在一种机制，能够从 G_t 中移除区块，以确保状态空间有限。我们定义了两种机制，将区块图中的区块数量限制在最多 M 个。概括而言，

(1) 当绝大多数代理都打算向 G_t 的某个子图追加内容时，图将被**剪枝**，使该子图成为新的全局区块图；(2) 若不满足剪枝条件，且 G_t 中的区块数量将超过 M，则状态**重置**，即以概率 1 转移至初始化状态。

更确切地说，对于区块图 G_t 和区块 $x \in G_t$，令 $\Gamma(G_t,\ x)$ 表示 G_t 的诱导子图，其中仅包含 x 及其后代节点。若存在任一区块 x，使得向 $\Gamma(G_t,\ x)$ 中某个区块追加内容的累积概率至少为 $1-\epsilon$，其中 $0 \le \epsilon < 0.5$，则原地对 G_t 进行剪枝，使新的区块图变为 $G_t = \Gamma(G_t,\ x)$，随后状态从$\curvearrowright G_t$转移。若有多个子图满足此条件，则选择其中最小的子图。因此，状态 S_t 中区块 x 的剪枝条件定义如下：

$$\Psi_{\overline{\pi}}(S_t, x, a_t, \mu) = \begin{cases} 1 & \frac{1}{N}\left(\mathbb{1}_{\Gamma(G_t,x)}(a_t) + (N-1)\sum_l \mu(l, G_t)\,\mathbb{1}_{\Gamma(G_t,x)}(\overline{\pi}(l))\right) > 1-\epsilon \\ 0 & \text{else,} \end{cases} \tag{5}$$

其中，a_t是代表性智能体的动作，ϵ是剪枝阈值，而 $\mathbb{1}_{\Gamma(G_t,\ x)}(a_t)$ 是指示函数：当动作 a_t 指定的区块位于 $\Gamma(G_t,\ x)$ 中时，该函数的值为 1。计算状态转移时，会先于任何转移概率检查剪枝条件

[4]通过为区块图定义不同的同构类，我们的模型可以扩展到分析其他区块链结构，例如基于有向无环图的区块链。

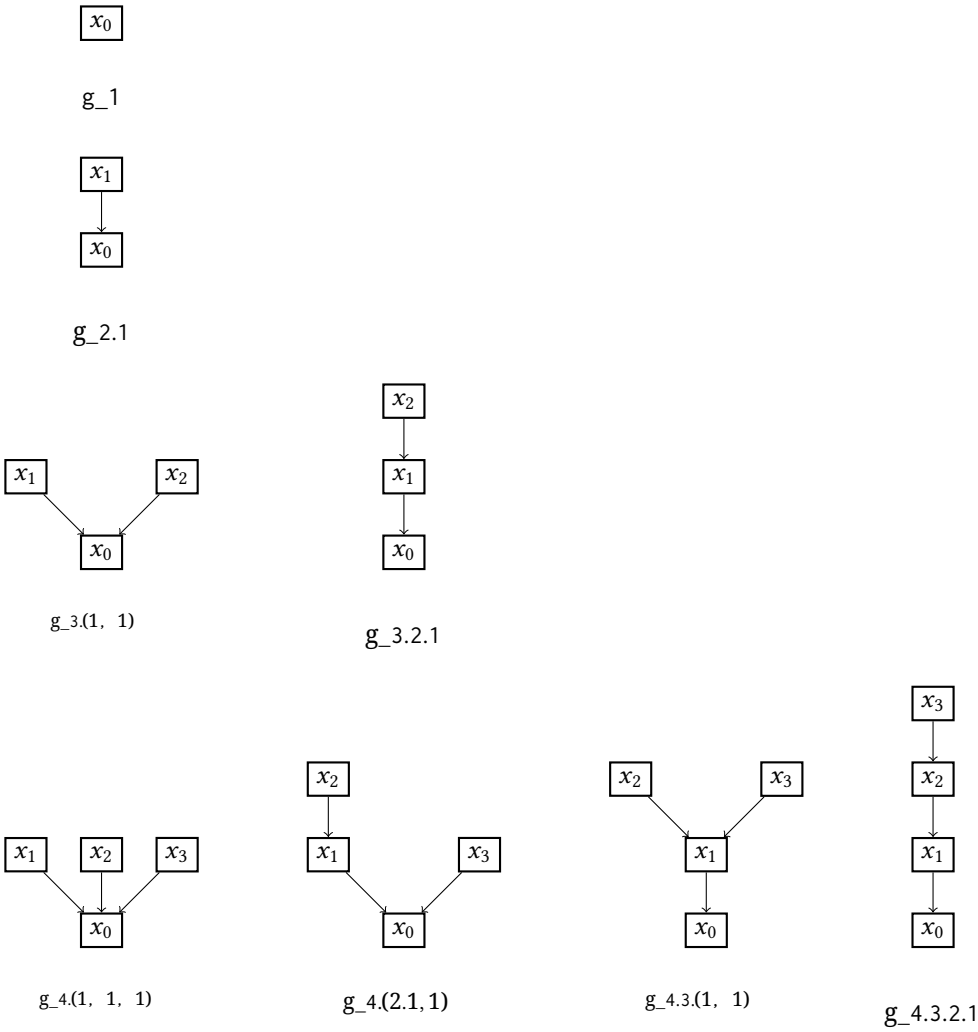


图 2。该图展示了由 1、2、3 和 4 个区块构成的中本聪图的所有同构类。每个区块状态下方都标注了该区块状态，其中包含一种嵌套编码，有助于将集合扩展到 5+ 个区块。

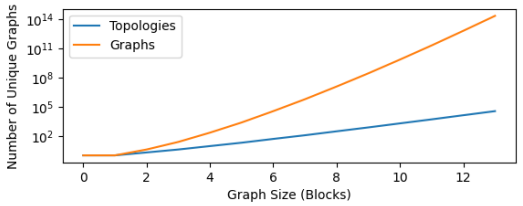


图 3。图中唯一图的数量与唯一图拓扑的数量随图中区块数量变化的关系。请注意，纵轴采用对数刻度。

计算得出，然后对于区块集合

$$\mathcal{X} = \{x : \Psi_{\overline{\pi}}(S_t, x, a_t, \mu) = 1\},$$

G_t 被替换为 $G_t = \Gamma(G_t, \widehat{x})$，其中

$$\widehat{x} = \min_{x \in \mathcal{X}} |\Gamma(G_t, x)|.$$

请注意，由于 $\epsilon < 0.5, \widehat{x}$ 是 $|\Gamma(G_t, x)|$ 的唯一最小值，因为 $\{\Gamma(G_t, x) : x \in \mathbf{X}\}$ 是 G_t 的嵌套子图集合，其中每个子图都是一组区块，代理“同意”在 π 下追加这些区块。

理想情况下，剪枝应频繁发生，这意味着代理通常能够就某个区块或子图达成一致。然而，许多局部策略和系统状态并非如此。因此，必须有一种额外机制来确保区块图保持有限：如果全局区块图包含 M 个区块且不满足剪枝条件，则状态**重置**为初始化状态的概率为 1。重置是未能就某个区块或子图达成一致的结果，因此，从系统设计者的角度来看，应避免采用会导致频繁重置的协议或区块选择策略。

5.2 转移概率

回顾一下，T_π 可以分解为各分量概率。下面说明如何计算每个概率。

5.2.1 全局区块图转移概率。从 G_t 转移到 G_{t+1} 的概率定义如下。令 $\Phi(G_t, G_{t+1})$ 为 G_t 中可能为空的区块集合，使得将一个区块追加到任意 $x \in \Phi(G_t, G_{t+1})$ 后，得到的区块图为 G_{t+1}。例如，参见图 2，

$$\Phi(\mathsf{g_3.(1,1)}, \mathsf{g_4.(2.1,1)}) = \{x_1, x_2\},$$

因为向 g_3.(1，1) 中的 x_1 或 x_2 追加区块，都会得到图拓扑 g_4.(2。1，1)。另一方面，

$$\Phi(\mathsf{g_3.(1,1)}, \mathsf{g_4.3.2.1}) = \varnothing,$$

因为在 g_3.(1，1) 中，没有任何一个区块可以被追加，并使结果得到 g_4.3.2.1 的图。

因此，只要区块图未被剪枝，且 G_{t+1} 中的区块数量不超过M，从 G_t 到 G_{t+1} 的转移概率为

$$P(G_{t+1} \mid S, a_t, \mu, \overline{\pi}) = \frac{1}{N}\left(\mathbb{1}_{\Phi(G_t, G_{t+1})}(a) + (N-1)\sum_l \mu(l, G_t)\,\mathbb{1}_{\Phi(G_t, G_{t+1})}(\overline{\pi}(l))\right). \tag{6}$$

从计算角度而言，Φ 可以通过结合动态规划算法和相当直接的图逻辑来高效计算；该算法利用了中本聪图的嵌套结构。

5.2.2 所有区块状态转移。接下来，我们考虑概率 $P(O_{t+1} \mid S_t, a_t, \mu, \pi, G_{t+1})$。在每次状态转移中，都会向全局区块图追加且仅追加一个新区块 x'（除非状态重置，此时转移概率已定义）。因此，O_{t+1} 可能取两个值。如果代表性智能体生成了 x'，那么 $O_{t+1} = O_{t+1}^{(1)}$，其特征条件为

$$\forall x \in G_t : O_{t+1}^{(1)}[x] = O_t[x]$$

and

$$O_{t+1}^{(1)}[x'] = 1.$$

如果代理没有生成 x'，那么 $O_{t+1} = O_{t+1}^{(0)}$，其中

$$\forall x \in G_t : O_{t+1}^{(0)}[x] = O_t[x]$$

and

$$O_{t+1}^{(0)}[x'] = 0.$$

代表性智能体生成 x' 的概率为

$$P(O_{t+1}[x'] = 1 \mid s, a_t, \mu, \overline{\pi}) = \frac{\mathbb{1}_{\Phi(G,G_{t+1})}(a_t)}{\left(\mathbb{1}_{\Phi(G,G_{t+1})}(a_t) + (N-1)\sum_l \mu(l)\mathbb{1}_{\Phi(G,G_{t+1})}(\overline{\pi}(l))\right)}. \tag{7}$$

其含义如下：如果 RA 和 $m-1$ 其他代理尝试在 $\Phi(G, G_{t+1})$ 中追加区块，代理 j 以概率 $\frac{1}{m}$ 生成 x'。如果 RA 没有尝试在 $\Phi(G, G_{t+1})$ 中追加区块，那么他们就无法生成 x'。

综上所述，

$$P(O_{t+1} \mid S_t, a_t, \mu, \overline{\pi}, G_{t+1}) = \begin{cases} P(O_{t+1}[x'] = 1 \mid S_t, a_t, \mu, \overline{\pi}) & \text{if } O_{t+1} = O_{t+1}^{(1)} \\ 1 - P(O_{t+1}[x'] = 1 \mid S_t, a_t, \mu, \overline{\pi}) & \text{if } O_{t+1} = O_{t+1}^{(0)}. \end{cases} \tag{8}$$

5.2.3 已接收区块向量转移概率。 G_t 中的区块按照独立过程传播，因此可以逐个区块计算 $P(B_{t+1} \mid S_t, a_t, \mu, \pi, G_{t+1}, O_{t+1})$。设 $X_{G_{t+1}}$ 为 $\overline{G_{t+1}}$中的区块集合。对于代理生成或在状态 S_t时已经接收的任意区块 x， $B_{t+1}[x] = 1$的概率为 1。对于任何其他满足 $B_t[x'] = 0$和 $O_t[x'] = 0$的区块 $x' \in X_{G_t}$，

$$P(B_{t+1}[x'] = 1 \mid S_t, a_t, \mu, \overline{\pi}, G_{t+1}, O_{t+1}) = P(H_{\alpha,\delta} = 1).$$

那么，如果 $z_1 = |\{x \mid x\colon B_t[x] = 0,\ B_{t+1}[x] = 1\}|$ 是代理在状态从 S_t 转移到 S_{t+1} 时获得的区块总数，而 $z_0 = |\{x \mid x\colon B_t[x] = 0,\ B_{t+1}[x] = 0\}|$ 是它未获得的区块数，

$$P(B_{t+1} \mid S_t, a_t, \mu, \overline{\pi}, G_{t+1}, O_{t+1}) = P(H_{\alpha,\delta} = 1)^{z_1}(1 - P(H_{\alpha,\delta} = 1))^{z_0}. \tag{9}$$

5.3 奖励

回顾一下，奖励函数的形式为

$$R_{\overline{\pi}}(S_t, a_t, S_{t+1}, \mu) = rQ(S_t, a_t, S_{t+1}, \mu, \overline{\pi}),$$

其中， $Q(S_t,\ a_t,\ S_{t+1},\ \mu,\ \pi)$ 是代理应获得奖励的关键路径区块数量。在大多数加密货币中，只有在以较高概率确定某个区块是否位于关键路径上之后，代理才能花费该区块的奖励。这与我们的剪枝条件恰好吻合：当某个区块几乎不可能再有新区块接在其后时，就会将其从区块图中移除。令 $X(G_t,\ x) \subset G_t$ 表示 G_t 中作为 x 祖先的区块集合。换言之，$X(G_t,\ x)$ 是在 x 时刻对 G_t 进行剪枝时，从 G_t 中移除的关键路径区块集合。于是，如果将区块图 G_t 剪枝为 $G_t = \Gamma(G,\ x')$，

$$Q(S_t, a_t, S_{t+1}, \mu, \overline{\pi}) = |\{x \in X(G_t, x') : O_t[x] = 1\}|. \tag{10}$$

如果达到区块上限 M 时状态重置，且没有进行剪枝，那么将不会发放奖励，因为 G_t 没有被剪枝。这是有意为之的，因为这会鼓励代理选择尽可能避免重置的策略。

5.4 计算平均场分布

对于任意选择的 μ，都可以求解 $\mathbf{M}_\pi$；但要使解成为平均场均衡，μ 必须与区块传播动力学保持一致，并且在所有代理（包括代表性智能体）之间保持对称。设分布 μ_π 表示非代表性智能体采用局部策略 π 时，各代理的局部区块图长期平均分布。由于系统动态复杂且状态空间规模庞大，解析求解 μ_π 颇具挑战。不过，由于系统中的所有代理都按照相同的过程接收区块，我们可以根据 $\mathbf{M}_\pi$ 的平稳分布计算 μ_π。在稳态下，代表性智能体的局部区块图分布与非代表性智能体相同。

具体而言，令 $f(l, G_t, \overline{\pi}, \mu)$ 表示全局区块图为 G_t 时，代表性智能体的局部区块图 l 的稳态概率。那么，当且仅当 μ_π 是 f 的不动点时，均值场分布 μ_π 是平稳的；也就是说，对每个 $t, G_t \in \mathbf{G}$，

$$\mu_{\overline{\pi}}(l, G_t) = f(l, G_t, \overline{\pi}, \mu_{\overline{\pi}}).$$

因此，计算 μ_π 可归结为计算 f 的不动点，这可以通过以下迭代算法完成。

首先，确定初始估计 μ_π^0。估计 μ_π^0 有多种方法，但在我们的实验中采用以下估计方法。对于任意全局区块图 G_t 和某个（非代表性）代理 i，令

$$p_D(G_t, x) = P\big(H_{\alpha,\delta} \le D(G_t, x) + 1\big),$$

and

$$p_C(G_t, x) = P\big(H_{\alpha,\delta} \le C(G_t, x) + D(G_t, x) + 1\big),$$

其中，$D(G_t, x)$ 是 G_t 中 x 的后代节点数量，而 $C(G_t, x)$ 是 G_t 中既非 x 的后代节点也非其祖先的区块数量。那么，某个区块 $x \in G_t$ 属于代理 i 的局部区块图的概率上界为

$$p_D(G_t, x) \le P(x \in L^i \mid G_t) \le p_C(G_t, x).$$

这是因为 $P(x \in L^i \mid G_t)$ 取决于自 x 生成以来已经过去的区块步数量，而根据 G_t 的构建顺序，可以通过相当直接的推理对该数量进行界定。$D(G_t, x)$ 是在 x 之后必定已经生成的区块数量，因为它们是 x 的后代节点。另一方面，$C(G_t, x)$ 是 G_t 中相对于 x 无法确定生成顺序的区块数量，因为它们可能在 x 之前或之后生成。然后，对于每个可能的局部区块图 l 和全局区块图 G_t，我们设

$$\mu_{\overline{\pi}}^0(l, G_t) = \prod_{x \in l} \frac{1}{2}\left(p_D(G_t, x) + p_C(G_t, x)\right),$$

也就是说，每个局部区块图的概率被初始化为上下界的中间值。

确定初始估计 μ_π^0 后，下一步是求解 $\mathbf{M}^{\mu_\pi^0}$ 并确定所得的稳态分布。对于所有 l，$G_t \in \mathbf{G}$，

$$\mu_{\overline{\pi}}^1(l, G_t) = f(l, G_t, \overline{\pi}, \mu_{\overline{\pi}}^0).$$

随后，计算 $f(l, G_t, \overline{\pi}, \mu_\pi^1)$，并重复这一过程，直到 f 达到不动点，此时便确定了 μ_π。

5.5 计算平均场均衡

回顾第4.6节，M 的平均场均衡由贝尔曼方程的不动点刻画。

$$\overline{\pi}_j^*(L_t^j) = \arg\max_a \sum_s T_{\overline{\pi}^*}(S_t, a, s, \mu_{\overline{\pi}})[R_{\overline{\pi}^*}(S_t, a, s, \mu_{\overline{\pi}}) + \gamma V^*(s)]$$

and

$$V^*(s) = \max_a \sum_{s'} T_{\overline{\pi}^*}(s, a, s', \mu_{\overline{\pi}})[R_{\overline{\pi}^*}(s, a, s', \mu_{\overline{\pi}}) + \gamma V^*(s')].$$

然而，这样的 $\pi_j^*(L_t^j)$ 并不一定存在，因为可能有多个状态对应于代表性智能体的同一个局部区块图，而每个状态的最优动作可能并不相同。如果与 S_t 相同的局部区块图对应的所有状态构成的集合是 $\sigma(S_t)$，那么必须对 π_j^* 施加额外约束，使其满足

$$\forall s, s' \in \sigma(S_t) : \overline{\pi}_j^*(s) = \overline{\pi}_j^*(s').$$

我们介绍两种在此约束下计算 π_j^* 的逻辑方法。

方法1。第一种方法是修改 $\pi_j^*(L_t^j)$ 的定义，以满足这一额外约束。代表性智能体不再针对当前状态选择能使期望未来奖励最大化的动作，而是针对所有可能引发局部区块图 L_t^j 的状态，最大化其期望未来奖励的某个函数（例如 L1 范数）。这意味着要求解

$$\overline{\pi}_j^*(L_t^j) = \arg\max_a \Big\| \sum_s T_{\overline{\pi}^*}(\widehat{s}, a, s, \mu_{\overline{\pi}})[R_{\overline{\pi}^*}(\widehat{s}, a, s, \mu_{\overline{\pi}}) + \gamma V^*(s)] \Big\|_{\widehat{s} \in \sigma(S_t)}$$

and

$$V^*(s) = \max_a \Big\| \sum_{s'} T_{\overline{\pi}^*}(\widehat{s}, a, s', \mu_{\overline{\pi}})[R_{\overline{\pi}^*}(\widehat{s}, a, s', \mu_{\overline{\pi}}) + \gamma V^*(s')] \Big\|_{\widehat{s} \in \sigma(s)}.$$

这样一来，代表性智能体就能保证对每个状态 $\widehat{s} \in \sigma(S_t)$ 都选择相同的动作。这种方法背后的直觉是，代理意识到有多个系统状态可能会产生它所观察到的局部区块图，因此，它不会针对某个特定状态选择最优动作，而是针对每个可能状态上的某种分布进行优化（对于 L1 范数而言，该分布是均匀分布）。

方法2。第二种方法是将 M_{π^*} 视为完全可观测的马尔可夫决策过程来求解，然后将所得的（完整）策略转换为局部策略。这只需先求解

$$\pi_j^*(S_t) = \arg\max_a \sum_s T_{\overline{\pi}^*}(S_t, a, s, \mu_{\overline{\pi}^*})[R_{\overline{\pi}^*}(S_t, a, s, \mu_{\overline{\pi}^*}) + \gamma V^*(s)]$$

and

$$V^*(s) = \max_a \sum_{s'} T_{\overline{\pi}^*}(s, a, s', \mu_{\overline{\pi}^*})[R_{\overline{\pi}^*}(s, a, s', \mu_{\overline{\pi}^*}) + \gamma V^*(s')]$$

然后使用以下方法从 π_j^* 推导出 π_j^*。在时间 t，当代表性智能体具有局部区块图 L_t^j 时，设 S' 为满足 $G' = L_t^j$ 和 $\forall x \in G'$ 的状态：$R'[x] = 1$。然后可以通过以下方式从完整策略推导出局部策略：

$$\overline{\pi}_j(L_t^j) = \pi_j(S').$$

换句话说，在局部区块图 L_t^j 和局部策略下，动作智能体 j 所采取的动作，就是在状态为 S' 时、全局区块图为 L_t^j 且该代理已收到所有区块的情况下，按照完整策略所采取的动作。直观而言，这反映了这样一种实际观念：

参数	描述	默认值
N	代理数量。	1000
M	全局区块中的最大区块数量 图。	5
α	单个时间步内生成区块的概率。 时间步。	0.001
δ	对于代理尚未收到的任何区块， 其在下一个时间步收到该区块的概率为 δ.	0.01
γ	未来奖励的折扣率。	0.99
ϵ	剪枝阈值为 $1-\epsilon$，这意味着如果某个概率 子图将被追加到 G_t 的概率大于 $1-\epsilon$，则会立即对区块图进行剪枝。	0.01
r	代表性智能体因生成 关键路径区块而获得的奖励。	1
$\overline{\pi}_0$	非代表性智能体的初始策略。	LCR

表1. 系统参数汇总。

加密货币矿工在选择动作时，会假定自己观察到的局部区块图就是 全局区块图。

实际上，大多数加密货币矿工所遵循的策略与方法2更为接近，因为挖矿软件通常在决定追加哪个区块时，完全不会考虑未接收区块。因此，在我们的实验中，我们采用方法2，并将方法1的探索留待未来研究。

6 实现细节

我们使用 Python3 实现了这一框架，以考察加密货币矿工的策略行为，并分析由此产生的区块图的结构与演化。为了直观展示各系统参数的实际影响，我们首先对 M 进行探索性分析。随后，我们介绍在实现中如何定义策略。

6.1 系统参数

我们的框架以若干参数作为输入，这些参数对应不同的系统动态。这些参数汇总于表1中；除非另有说明，我们在实验中均使用默认值。为了提供有助于理解或扩展本研究的见解，我们简要讨论每个参数对矿工行为和系统动态的影响。

6.1.1 智能体数量。在平均场近似下，假设代理数量足够多，以至于每个代理对状态的影响都很小。一般而言，代理数量影响均衡策略的唯一方式，是通过代表性智能体生成每个区块的概率（由于假设代理的能力相同，该概率为 $1/N$）。因此，N 只会影响代理所获得奖励的大小，而不会影响其动作选择，所以我们的所有实验均使用默认值 $N = 1000$。

6.1.2 最大区块图规模。 使用我们的实现，我们能够在合理的（墙上时钟）时间内，求解全局区块图最多包含 $M = 7$ 个区块的马尔可夫决策过程，这

对应于 54053 个唯一状态。然而，对于包含 4、5、6 和 7 个区块的区块图，我们观察到几乎相同的行为，因此我们的大多数实验都采用最大区块图规模为 5，这对应于 1537 个状态。在实践中，矿工很少需要在超过 5 个区块之间进行选择，因此 $M = 5$ 是一个合理的默认参数。

6.1.3 区块生成与传播。单独来看，参数 α 和 δ 可以任意设定，因为时间步的持续时间并不固定。选择 $\alpha = 0.001$ 可以对应于每纳秒生成一个区块，也可以对应于每世纪生成一个区块，具体取决于时间步的持续时间。然而，将（α，δ）作为一对参数来选择则重要得多，因为它对应于区块生成和传播的相对速率。在我们的实验中，比例为 $\frac{\delta}{\alpha} = 10$时会得到稳定的结果，尽管这并非唯一能得到稳定结果的比例。例如，这可能对应于 10 分钟的预期区块时间和 1 分钟的预期传播延迟。

6.1.4 折扣因子。 γ 的选择不会显著影响我们观察到的均衡。在实践中，折扣因子应假定相当接近 1，因为区块之间的时间通常足够短，以至于矿工对收到一个区块奖励的估值，与对收到下一个区块奖励的估值相同。

6.1.5 剪枝阈值。一般而言， ϵ 的选择对应于在判定某个区块是否位于关键路径上之前所要求的确定性程度。实际上，一旦某个区块从全局区块图中被剪枝，就可以认为该区块已最终确定，因为它无法再被追加到区块图中，也无法从区块图中移除。因此，在研究中本聪式加密货币的结算和最终性时，剪枝阈值的选择或许值得单独探讨。不过，我们将对 ϵ 的分析留待未来工作，并在实验中采用 $\epsilon = 0.01$ 作为默认值。这意味着，任何被追加到的概率低于 0.01 的区块都会从区块图中被剪枝。

6.1.6 奖励。只要奖励为正，代表性智能体生成区块所获得奖励的大小就没有实际意义，因此为方便起见，我们在所有实验中都设定 $r = 1$ 。

6.1.7 初始策略。我们观察到的均衡很大程度上取决于 NRA 遵循的初始策略。这是因为只有在区块图被剪枝时，代表性智能体才能获得奖励。对于不会导致剪枝的初始策略，代表性智能体无论遵循何种策略，都无法获得正奖励。对于会导致剪枝的初始策略，代表性智能体会尝试选择能最大化以下两项的动作：i）发生剪枝的概率，以及 ii）其生成关键路径区块的概率。这与实际情况的直觉十分吻合——如果所有矿工都只是随机选择一个区块进行追加，那么形成有意义的关键路径的可能性就很低，因此 RA 采取什么行动并不重要。对于能够形成有意义关键路径的策略（例如 LCR），RA 最好遵循相同的策略。

6.2 实现策略

完整策略是从状态到动作的映射，而局部策略是从局部区块图到动作的映射。每个动作都对应于代理局部区块图中的一个特定区块。虽然这一定义在理论上已经足够，但在实现此类策略时会出现实际问题。下面我们将讨论这些问题。

6.2.1 局部区块图为空的代理。回想一下，尚未收到 G_t 的根区块的代理，其局部区块图为空。此类代理没有可用动作，因此无法向 G_t追加区块。这会影响从 G_t转移到

G_{t+1}的概率，因为在时间 t，只有一部分 N代理能够生成区块。为了解决这一问题，我们将方程 6 乘以一个因子 κ进行归一化，其中 $\frac{1}{\kappa}$ 表示局部区块图中没有区块的代理所占的比例（可能包括代表性智能体）。

6.2.2 拓扑相同的区块。在实现确定性策略时还会遇到另一个问题——必须有一种确定性的方法来选择拓扑相同的区块，因为交换这些区块在图中的位置并不会改变图的拓扑结构。例如，考虑图2中的图 g_4.(1，1，1)。区块 x_1，x_2 和 x_3 在拓扑结构上都相同，因为交换它们的位置不会改变图的拓扑结构。为了确定马尔可夫决策过程的不动点，必须确保局部区块图相同的任何代理都会选择同一个区块。

为了解决这个问题，我们假设，尽管区块在功能上完全相同，但仍然可以彼此区分。然后，所有代理都遵循一条确定性规则，在拓扑相同的区块之间进行选择，例如始终选择区块哈希值最低的区块，或者在我们的可视化示例中，始终选择“最左侧”的区块。

7 结果

为了展示我们框架的实用性，我们用它证明了两个主要结果。首先，我们研究区块传播延迟变化对 PoW效率的影响——即最终进入关键路径的区块所占的比例。我们的观测与描述网络延迟和 PoW效率之间关系的理论结果高度吻合。这种吻合验证了我们使用平均场近似的合理性，因为我们框架的平稳行为与预期的理论结果一致。

接下来，我们考察 M 的平均场均衡。通过穷举搜索证明，我们发现，对于中本聪式加密货币，LCR 是最优区块选择策略，因为它在所有均衡策略中具有最高的平稳 PoW效率。然而，我们发现最长链规则相当脆弱，因为只有当非代表性智能体的初始策略与 LCR 相当接近时，我们的实现所确定的均衡才会收敛到 LCR。

7.1 网络延迟与 PoW效率

在异步区块链协议中，将区块传播到网络所需的时间（网络延迟）会显著影响协议的 *PoW*效率 ——即最终进入关键路径的区块所占的比例。这是因为随着网络延迟增加，代理在不一致的局部区块图上进行挖矿的时间也会增加，从而使产生非关键路径区块的可能性更高。理论结果表明，遵循 LCR 的区块链的 PoW效率为

$$\frac{1}{1+\alpha\Delta},$$

其中，Δ 是端到端网络延迟 [8, 21]。这一理论关系可用于检验我们所采用的平均场近似在多大程度上能够复现实际系统行为。为此，我们旨在证明以下命题。

命题1。 在平均场近似下，遵循 *LCR* 的区块链的 *PoW*效率为

$$\frac{1}{1+\alpha\Delta},$$

其中，α 是区块生成率，Δ 是网络延迟。

为检验这一说法，我们对除 δ之外的每个参数都采用表1中所述的默认值，并改变 δ以研究网络延迟对 PoW效率的影响。对于每个

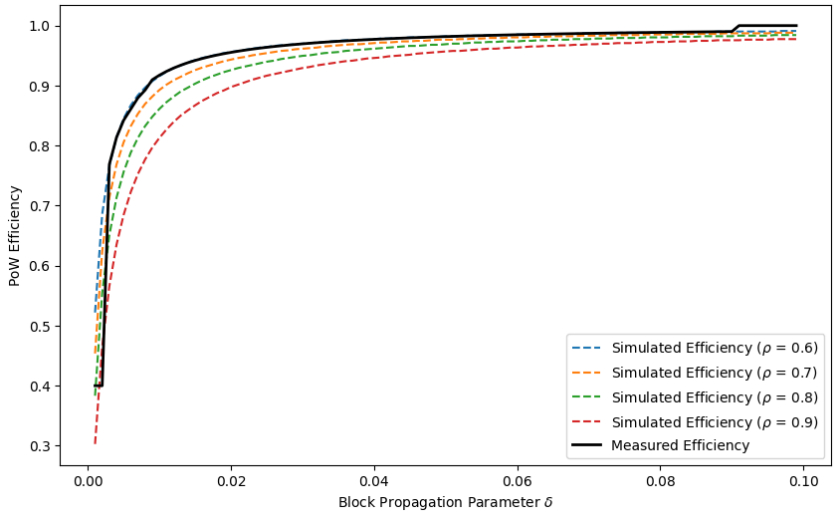


图4。非关键路径区块向区块图追加的速率与区块传播参数 δ的关系。虚线表示在不同网络延迟选择下的理论挖矿效率，实线表示我们在均衡状态下测得的挖矿效率。

δ的取值，网络延迟 Δ 可计算为使 ρ比例的代理收到区块所需的时间步数。请注意，在我们的模型中，要保证所有代理都收到区块所需的时间步数是无限的，因此我们针对几个对应于有限网络延迟的 $\rho < 1$ 取值，对 PoW效率进行建模。

对于 ρ的每个值，我们将理论PoW效率与LCR均衡策略下M的平稳PoW效率进行对比绘图。如图4所示，M在均衡状态下的PoW效率与预期结果高度一致，从而证明了命题1。

7.2 均衡最优性

我们发现，LCR并不是M唯一的平均场均衡。事实上，由我们的框架确定的均衡极易受到初始策略 π_0变化的影响。一般而言，不会导致频繁剪枝的初始策略会产生具有相同特征的均衡。这是因为在不存在剪枝的情况下，代表性智能体无法通过选择任何策略获得非零奖励，因此对于几乎所有状态下的动作都无差异。在这种情况下，所有可能的策略都是弱优势策略，因为无论代表性智能体采取何种动作，其获得的奖励都为零。

鉴于 M 存在多个均场均衡，均衡质量的问题便随之而来。比较均衡有许多可能的指标，但我们特别关注 PoW效率，因为它与 PoW区块链中的安全性、共识稳定性和能源效率直接相关 [8, 21]。直觉上，在中本聪式区块链中，LCR 应当是能使 PoW效率达到最优的区块选择策略，因为：i）奖励通常只发放给最终位于关键路径上的区块；ii）生成位于关键路径上的区块，一种合理的启发式方法是将其附加到当前关键路径上。比特币和其他中本聪式加密货币多年来一直采用 LCR，这也提供了一定程度的隐含经验依据，支持这一结论。然而，据我们所知，LCR 的最优性

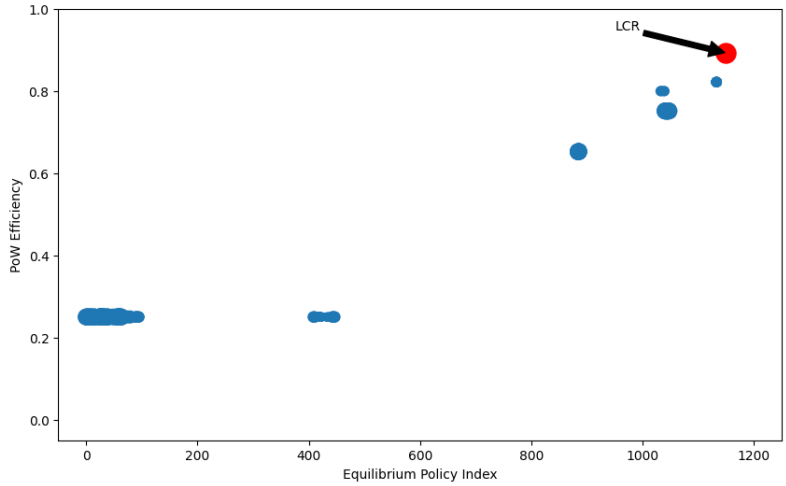


图 5. M 的每个均衡策略对应的 PoW效率。横轴表示均衡策略的索引，可忽略。每个标记的大小对应于遇到该均衡策略的频率。

作为区块选择策略尚未得到证明。我们的框架支持这样的证明。具体而言，我们旨在证明以下论断。

断言 2. 在第 *4* 节的假设下，*LCR* 是使诱导马尔可夫链的平稳*PoW*效率最大化的平均场均衡策略。

我们通过穷举搜索证明这一断言。也就是说，我们计算每一种可能的平均场均衡，并计算稳态下由此产生的 PoW效率。为此，我们将区块图的大小限制为 4 个区块[5]，，这对应于 1152 种唯一策略。在这些策略中，有 150 种是 M 的平均场均衡，尽管其中大多数均衡是弱支配的，并且会使代表性智能体获得零奖励。

图 5 展示了每种可能的均衡策略对应的 PoW效率，标记的大小表示遇到该均衡的频率。我们发现，LCR 作为均衡策略是唯一最优的；与 LCR 越相似的策略，其 PoW效率越高，而与 LCR 差异显著的策略则不然。由于不存在任何可能的均衡策略能够超越 LCR 的 PoW效率，这便构成了对断言 2 的证明。

8 未来工作与结论

我们指出了若干有趣的方向，可据此扩展 M，以支持不同类型的分析；随后给出几点总结性说明。

8.1 未来工作

其他奖励函数。 奖励函数在塑造矿工激励方面发挥着根本作用。我们的框架有助于研究奖励结构的变化（例如，奖励创建非关键路径区块或提交部分份额的代理）如何影响均衡策略和

[5]唯一策略的数量随着区块图规模的增大而超指数增长。对于规模为 5 的最大区块图，共有 165,888,000 个唯一策略，因此，当区块规模大于 4 时，穷举搜索并不切实际。

系统动态。理论上，包含足够高交易费用的区块可能会导致共识过程中出现低效率；在我们的框架下，可以通过添加状态变量来考察这一问题，这些变量用于指示区块的费用是“高”“低”还是介于两者之间。

挖矿算力的非均匀分布。 在我们的模型中，所有代理的挖矿算力都相同。这种对称性使我们能够针对任意一组系统参数求解平均场均衡。在现实世界的加密货币系统中，参与者的挖矿算力分布并不均匀，而且只有当攻击者的挖矿算力达到总算力的某个阈值后，多种攻击才具有可行性。只需修改转移概率函数，我们的框架便能够处理代表性智能体占有更大挖矿算力份额的情况。在这种情况下，我们的框架可以计算代表性智能体针对任意NRA策略的最佳响应；但由于代理不再具有对称性，因此无法搜索均衡策略。

扣留攻击。 扣留攻击是指矿工扣留区块以获取战略优势的策略。最常见的两种扣留攻击是自私挖矿和私有双花。在我们的框架中，可以通过允许代表性智能体在满足特定条件之前扣留区块来模拟这些攻击。静态扣留攻击（不会根据不断变化的系统动态进行调整）只需对状态空间进行很少的修改，甚至无需修改即可模拟，因为可以将扣留策略和收益纳入转移概率函数和奖励函数中。

替代区块链结构。 尽管我们的框架基于中本聪图，但基于其他图结构的区块链，例如有向无环图（DAG）或并行链，也备受关注。要实现这类结构，需要修改状态空间和转移概率函数，以适应不同的图拓扑；而系统动态和求解方法则保持不变。

替代区块传播模型。在许多区块链系统中，不同类型的代理接收区块的速率可能各不相同，这取决于网络拓扑、地理距离或带宽限制。同样，也有多种可用于传播区块的广播算法。要研究这些网络动态的影响，可以在模型中添加额外的平均场变量，并增加状态变量，以指明不同类型的区块应如何传播。

替代最优性条件 在第7节中，我们发现 LCR 在最大化 PoW 效率方面是最优的。不过，也可以通过其他指标来衡量区块选择策略的质量，例如结算时间、抗攻击性 [25],以及奖励分配的公平性等。事实上，大多数基于比特币构建的协议，恰恰是为了改善其中一项或多项指标而设计的。一般而言，我们的框架能够为每种均衡策略计算完整的转移概率矩阵，因此，对于任何通过对 M 的区块结构或转移概率函数进行充分修改即可表示的协议，都可以在稳态下对这些指标进行刻画。如果存在多个均衡策略，这就为确定任何协议在任何最优性条件下的最优区块选择策略提供了一种直接且可扩展的方法。

8.2 结论

区块链已成为去中心化系统的基础技术，但评估和提升其性能与安全性仍是一项复杂的挑战。在本研究中，我们提出了一个用于建模区块链动态的框架，通过求解一系列部分可观测马尔可夫决策过程，计算区块链增长博弈的均场均衡。由此，我们提供了一种结构化且可扩展的方法，用于分析共识机制和矿工行为。我们的

框架支持对不同的系统动态、奖励结构和区块选择策略进行严格评估，并揭示这些因素如何影响稳态下的区块链性能。

我们研究了网络延迟与PoW效率之间的权衡，发现我们在框架中观察到的平稳PoW效率与理论认识高度一致。这一结果验证了我们采用平均场近似和求解方法的合理性。此外，我们首次证明了LCR在中本聪式区块链中的最优性。通过穷举搜索，我们表明，在所有均衡策略中，LCR能够实现最高的平稳PoW效率，从而证明了其在这些系统中持续占据主导地位的合理性。尽管我们的研究聚焦于中本聪式区块链，但该模型也适用于其他区块链架构，使其成为未来研究提高区块链系统效率、安全性和性能的多用途工具。

参考文献

[1] Sachin Adlakha、Ramesh Johari 和 Gabriel Y Weintraub。2015。多玩家动态博弈的均衡：存在性、近似与市场结构。《经济理论杂志》156（2015），269–316。[2] Stefano V Albrecht、Filippos Christianos 和 Lukas Schäfer。2024。《多智能体强化学习：基础与现代方法》。麻省理工学院出版社。[3] Vitalik Buterin 等。2013。以太坊白皮书。GitHub 仓库 1，22-23（2013），5–7。[4] Peter E. Caines、Minyi Huang 和 Roland P. Malhamé。2018。《平均场博弈》。施普林格国际出版公司，Cham，345 372 10 1007 978 3 319 44374 4_7 – 。https://doi.org/ 。 / - - - - Christian Decker 和 [5] Roger Wattenhofer。 。比特币网络中的信息传播。载于《IEEE P2P 2013会议论文集》。IEEE，1–10。[6] Amir Dembo、Sreeram Kannan、Ertem Nusret Tas、David Tse、Pramod Viswanath、Xuechao Wang 和 Ofer Zeitouni。2020。一切皆为竞赛，而中本聪总能获胜。载于《 2020 年 ACM SIGSAC 计算机与通信安全大会论文集》。859–878。[7] Juan Garay、Aggelos Kiayias 和 Nikos Leonardos。2024。比特币骨干协议：分析与应用。《ACM学报》71，4（2024），1–49。[8] Yanni Georghiades、Karl Kreder、Jonathan Downing、Alan Orwick 和 Sriram Vishwanath。2022。通过分层最长链规则实现可扩展的多链协调。载于《2022年IEEE国际区块链会议（Blockchain）》。IEEE，468–475。[9] Yanni Georghiades、Rajesh Mishra、Karl Kreder 和 Sriram Vishwanath。2023。不需要多数：少数派对手发起私有双花攻击的理性分析。《分布式账本技术：研究与实践》（2023）。[10] Milos Hauskrecht。2000。部分可观测马尔可夫决策过程的价值函数近似。《人工智能研究杂志》13（2000），33–94。[11] Daira Hopwood、Sean Bowe、Taylor Hornby、Nathan Wilcox 等。2016。Zcash协议规范。GitHub：美国加利福尼亚州旧金山 4，220（2016），32。[12] Alejandro Jofré、Angel Pardo、David Salas、Victor Verdugo 和 José Verschae。2021。区块链挖矿博弈的收敛速率：马尔可夫方法。arXiv预印本 arXiv:2107.08077（2021）。[13] Aggelos Kiayias、Elias Koutsoupias、Maria Kyropoulou 和 Yiannis Tselekounis。2016。区块链挖矿博弈。载于《2016年ACM经济学与计算会议论文集》。365–382。[14] Mykel J Kochenderfer、Tim A Wheeler 和 Kyle H Wray。2022。《决策算法》。麻省理工学院出版社。[15] Vikram Krishnamurthy。2016。《部分可观测马尔可夫决策过程》。剑桥大学出版社。[16] Jean-Michel Lasry 和 Pierre-Louis Lions。2007。平均场博弈。《日本数学杂志》2，1（2007），229–260。[17] Zongxi Li、A Max Reppen 和 Ronnie Sircar。2024。加密货币挖矿的平均场博弈模型。《管理科学》70，4（2024），2188–2208。[18] Andrew Miller 和 Rob Jansen。2015。{Shadow-Bitcoin}：通过直接执行{多线程}应用程序实现可扩展仿真。载于第八届网络安全实验与测试研讨会（CSET 15）。[19] 中本聪。2008年。比特币：一种点对点电子现金系统。《去中心化商业评论》（2008），21260。[20] Kartik Nayak、Srijan Kumar、Andrew Miller 和 Elaine Shi。2016。顽固挖矿：自私挖矿的推广及其与日蚀攻击的结合。载于《2016年IEEE欧洲安全与隐私研讨会（EuroS&P）》。IEEE，305–320。[21] Rafael Pass、Lior Seeman 和 Abhi Shelat。2017。异步网络中的区块链协议分析。载于《密码技术理论与应用国际年会》。施普林格，643–673。[22] Martin L Puterman。2014。《马尔可夫决策过程：离散随机动态规划》。约翰·威立父子出版公司。

[23] Jeff Reed。2017。莱特币：莱特币加密货币与莱特币挖矿简介。[24] Lloyd S Shapley。1953。随机博弈。《美国国家科学院院刊》39，10（1953），1095–1100。[25] Ren Zhang 和 Bart Preneel。2019。制定通用指标：评估工作量证明共识协议的安全性。载于《*2019*年*IEEE*安全与隐私研讨会（*SP*）》。IEEE，175–192。