arXiv:2601.00523v1 [cs.GT] 2 Jan 2026

CoinAlg 困境：集体投资算法中的盈利能力与公平性的权衡

Andrés Fábrega*
康奈尔科技，*IC3*

James Austgen*
康奈尔科技，*IC3*

Samuel Breckenridge
康奈尔科技，*IC3*

Jay Yu
Pantera Capital

Amy Zhao
Ava Labs ，*IC3*

Sarah Allen
Flashbots ，*IC3*

Aditya Saraf
康奈尔科技，*IC3*

Ari Juels
康奈尔科技，*IC3*

摘要

集体投资算法（CoinAlgs）是日益流行的系统，它们为投资者社群部署共享交易策略。其目标是让复杂的投资工具（通常基于人工智能）普及化。

我们识别并展示了 CoinAlgs 中一种根本性的盈利能力—公平性权衡，并将其称为*CoinAlg* 困境：CoinAlgs 无法在不因套利而损失利润的情况下确保经济公平性。

我们提出了 CoinAlgs 的形式化模型，并定义了隐私（算法披露不完整）和经济公平性（对抗性内部人员进行价值提取）。我们证明了两个相辅相成的结果，共同揭示了 CoinAlg 困境。首先，CoinAlg 中的隐私是对经济公平性发起内部人员攻击的前提。反过来，在博弈论模型中，缺乏隐私（即透明性）会使套利者能够削弱 CoinAlg 的盈利能力。

我们利用去中心化交易所 Uniswap 的数据，对 CoinAlg 困境的两面进行了实证研究。我们量化了针对透明硬币算法进行套利的影响，并展示了私有 CoinAlg 带来的风险：即使是低带宽的隐蔽信道信息泄漏，也会导致不公平的价值提取。

1 引言

本文开创性地研究了我们所称的集体投资算法（*CoinAlgs*）中的权衡：这类算法由用户群体共享，用于指示或代其执行金融交易，以实现盈利性投资策略。

CoinAlgs 在传统金融领域早已存在。例如，BlackRock Systematic 或 Renaissance Technologies 等公司的量化策略会为众多客户投资组合执行交易。相关变体还包括 Wealthfront 等机器人顾问，它们使用共享的人工智能引擎进行财务规划。

CoinAlgs 近期也开始在传统金融机构以外获得关注。面向个人投资者的人工智能交易软件包已陆续推出，可用于加密货币和传统金融资产交易，例如 [26, 33]。在去中心化金融（DeFi）领域，ElizaOS（代币 AI16Z）作为“首批已知的由人工智能控制的风险投资基金之一”，其峰值市值达到 $2+ 十亿 [5, 7]，；而 Truth Terminal 是一个通过在 X 上发布推文推动社区交易的人工智能代理，它促成了 Goatseus Maximus（GOAT）的发行。这是一种峰值市值达 $1+ 十亿 [31, 34]的迷因币。

CoinAlgs 通常承诺让散户投资者也能获得机构级投资策略或专业交易情报。然而，新近出现的 CoinAlgs——面向消费者的直接交易工具和基于 AI 的加密货币平台——缺乏受监管中介机构所提供的投资者保护。因此我们提出问题：CoinAlgs 中存在哪些设计权衡与风险？

在本项工作中，我们证明了 CoinAlgs 的设计者和用户面临一个根本性困境。CoinAlg 的交易策略必须至少部分私有，否则就必须透明。然而，这两种选择都会使投资者的利润面临风险：

- 如果**私有**，CoinAlg 可能会向内部人员泄露信息（例如通过隐蔽信道），从而使其能够对投资者进行不公平且可能无法被发现的价值提取。
- 如果**透明**，CoinAlg 会公开泄露信息，使其暴露于套利之下，并导致利润下降——我们将其称为透明性成本。

我们将这种张力称为*CoinAlg* 困境：任何 CoinAlg 都必须牺牲经济公平性或盈利能力。

1.1 我们的研究成果

我们的工作首次对 CoinAlg 困境所体现的权衡进行了形式化处理，主要包含三项成果，分别涉及 CoinAlgs 的隐私、透明性和实证研究。

*这些作者对本项工作的贡献相同。

CoinAlg 的隐私与经济（不）公平性。 如果根据 CoinAlg 的公开视图得出的预期交易概率分布与 CoinAlg 实际执行的交易之间存在不可忽略的统计距离，我们就称该 CoinAlg 是私有 的。换言之，公开视图不足以预测其交易。我们在工作中给出了 CoinAlg 隐私的形式化定义。

我们还提出了一种定义，将公平性视为衡量价值提取的指标：对手利用内部知识来利用 CoinAlg 的交易（例如发动三明治攻击）。这一界定是去中心化金融中类似概念（例如"矿工可提取价值／最大可提取价值"（MEV））的内部人员变体 [1, 8, 27, 36]。

本文的主要结果（定理1）表明，经济上的不公平意味着隐私。直观来说，隐私是赋予对抗性内部人员优势的必要条件。（反之则不一定成立：隐私并不意味着对手掌握了可采取行动的内部信息。）

CoinAlgs 的透明性。 我们还探讨了非私有但透明的 CoinAlgs。公众能够访问透明的 CoinAlg（例如，其算法开源），从而可以预先了解其交易。我们证明，这会使 CoinAlg 面临对手$\mathcal{A}$套利的风险。实际上，我们的透明性模型仅依赖于一个较弱的假设：$\mathcal{A}$ 能够模拟 CoinAlg；它并不要求$\mathcal{A}$ 对 CoinAlg 进行逆向工程。

将$\mathcal{A}$ 与透明的 CoinAlg 之间的互动建模为最后通牒博弈后，我们证明$\mathcal{A}$ 可以采用冷酷触发策略——以使 CoinAlg 的交易失效相威胁 —— 迫使 CoinAlg 暴露于 $\mathcal{A}$ 的套利之下（定理*4*）。我们模型背后的思路本身也可能具有研究价值，因为其中涉及这样一种重复博弈：故意使交易失效构成触发策略的基础。

实验。 我们利用以太坊上去中心化交易所 Uniswap V3 超过一年的历史交易数据（2024年6月至2025年7月），模拟一种能够正确预测未来资产价格且可盈利的 CoinAlg。我们研究 CoinAlg 困境的两面。我们证明，即便私有 CoinAlg 仅有少量信息泄露，也足以实现显著的价值提取。反过来，我们也表明，即便采取适度的套利防御策略，也会产生显著的透明性成本，从而侵蚀利润。

虽然为便于阐述，我们聚焦于加密货币 CoinAlgs，但我们的基本结果具有更广泛的适用性。例如，这些结果强调了传统金融中的投资者保护为何对于防范私人 CoinAlgs 中的价值提取攻击至关重要。

1.2 贡献

总之，我们的工作开启了对 CoinAlgs 中盈利能力与公平性权衡的研究——我们将其称为 CoinAlg 绑定。在介绍 CoinAlgs 的背景（第2节）以及它们当前的应用现状后，我们作出以下贡献：

- **隐私与公平性：** 基于 CoinAlgs 的一般形式化模型（第3节），我们定义了 CoinAlgs 中的隐私和公平性，并证明隐私是不公平性乃至内部人价值攫取的前提（第4节）。
- **透明性：** 通过对 CoinAlg 和对抗性套利者进行博弈论建模，我们证明，缺乏隐私，即透明性，会降低 CoinAlg 盈利能力，造成透明性成本（第5节）。
- **实验：** 我们基于历史以太坊市场开展6节所述的实验，为我们关于隐私、公平性及透明性成本的分析结果提供了实证支持。
- **防护措施：** 我们认识到，尽管 CoinAlgs 存在缺陷，但它们已不可避免地成为投资领域的一部分，因此简要提出两项实用防护措施，或可为*CoinAlg*投资者提供启发式保护，避免在私有 CoinAlgs 中遭受价值提取（第7节）。

我们对相关工作的回顾（第8节）凸显了现有研究对 CoinAlg 困境缺乏探讨，也彰显了我们在 DeFi 环境下提出的博弈论模型的新颖性。我们在结论中（第9节）指出，CoinAlg 困境体现了集体投资算法设计中的内在局限：必须在隐私带来的公平性风险与透明性所面临的套利风险之间作出选择。

2 什么是 CoinAlgs?

CoinAlgs 有两个主要特征：

1. 集体性：CoinAlg 的用户受共同交易逻辑约束，该逻辑驱动投资。2. 算法性：CoinAlg 的交易逻辑在本质上是算法。

换言之，CoinAlgs 是驱动集体投资行为的算法。这两个定义性属性正是 CoinAlg 困境之所以重要的原因：集体性带来了信托义务，进而产生公平性义务；而算法性则带来了透明性或隐私方面的可能性。

CoinAlg类别。 许多看似毫不相关的金融服务——涵盖传统金融和 Web3——都属于 CoinAlgs 的范畴。其中主要包括：

传统金融中的算法交易。投资者资产的集体性管理是传统金融（又称 TradFi）中一种成熟的实践，在高频交易（HFT）公司、对冲基金和量化投资公司中均有应用。这些机构的平台通常明确属于 CoinAlgs，通过汇集

投资者的资产，并代表他们以算法方式执行交易。

其他传统金融 CoinAlgs 的例子则更为隐晦，涉及投资者共享交易算法，但各自独立行事。许多成熟的传统金融公司（如 Vanguard、Schwab 等）提供的机器人顾问便是一例，不过这些服务会根据个人目标进行定制。广受欢迎的社交交易平台 eToro 则允许投资者之间明确复制交易算法。

面向投资者的平台。面向投资者的算法交易平台——例如用于股票交易的 Composer 和 QuantConnect，以及用于加密货币交易的 Zignaly——使个人投资者能够开发、分享或授权使用交易算法，同时保留对其资产的独立控制权。由于这些平台允许投资者社群共享交易逻辑，因此它们属于 CoinAlgs。

人工智能驱动的投资型去中心化自治组织和社群。在 DeFi 领域，将人工智能驱动的投资算法与去中心化自治组织（DAO）结合起来引发了广泛关注；DAO 是投资者汇集资金的智能合约。这类 CoinAlgs 的著名例子包括 AI16Z 和 AI-XBT。Terminal of Truths 则是一种不那么显而易见的 CoinAlg，详见第 1节。

在这三类 CoinAlgs 中，传统金融中的 CoinAlgs 是托管式的，这意味着企业以经纪交易商的身份持有投资者的资金，并受投资者保护法规约束。面向投资者的 CoinAlgs 和去中心化金融 CoinAlgs 通常是非托管式的，因此不受传统投资者保护制度的约束。

我们对 CoinAlgs 的定义不包括许多受欢迎的金融服务。例如，传统金融中的指数基金并非 CoinAlgs：它们并非以算法方式运行，因此没有实现透明性的机会。

CoinAlgs 的属性。正如示例所示，CoinAlgs 在两个维度上可能有所不同。

首先是用户资产的管理方式。这种方式可以是集中式，即将资产汇集到一个中心位置（例如，由人工智能驱动的去中心化自治组织、量化投资公司等）；也可以是分散式，即分别交易各项资产（例如，机器人顾问）。

第二个维度是 CoinAlg 的交易算法的获取方式。这同样可以是集中式或分散式：前者指算法位于单一的中心位置（例如，机器人顾问）；后者则指算法的多个副本分散部署（例如，开源交易机器人）。

我们在图 1中展示了几个 CoinAlgs 示例，并按其资产管理方式和算法类型进行分类。请注意，右下象限为空；据我们所知，目前没有任何 CoinAlgs 会整合资产，却采用分散式算法。

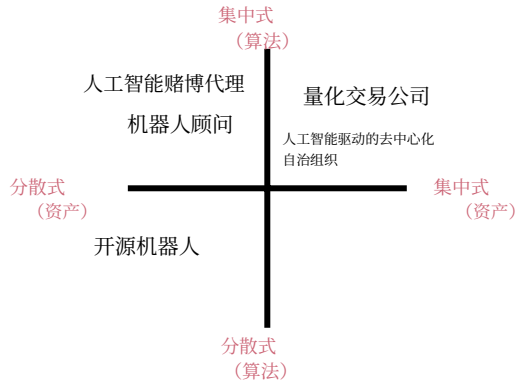

图 1：CoinAlgs 示例，以及用户资产和其交易算法是集中式还是分散式。

机器人顾问	AUM	客户数量	可见性
Vanguard	$206.6B	1.1M	私有
Schwab	$65.8B	262K	私有
Betterment	268亿美元	615K	私有
Wealthfront	214亿美元	307K	私有
Acorns	47亿美元	4.4M	私有

图2：热门机器人顾问概览，展示其管理资产规模（AUM）、个人客户数量，以及其是否为私有或透明。管理资产规模和客户数量的数据截至《福布斯》2024年8月发布的报告 [3]。

2.1 CoinAlgs 的实际应用

我们观察到 CoinAlgs 在实际应用中的行为呈现出三种模式，这些模式为我们的研究提供了依据。

模式 #1：备受关注。CoinAlgs 一直备受金融界关注。例如，在传统金融领域，机器人顾问已发展成为一个重要行业，由领先的金融服务公司（如 Vanguard、Wealthfront、嘉信理财等）提供服务，并管理着大量资产：根据《福布斯》2024年8月的一份报告，按资产管理规模计，前10名机器人顾问管理的资产超过3500亿美元，服务的个人客户超过670万名 [3]。参见图2，其中概述了几家颇受欢迎的机器人顾问，它们均位列资产管理规模前10名。

去中心化金融 CoinAlgs——尤其是人工智能驱动的去中心化自治组织——也备受关注，并于2025年初迅速走红，开始在 Web3 讨论中占据重要位置。然而，与机器人顾问不同，目前尚无反映这一行业规模的人工智能驱动的去中心化自治组织全景调查。因此，我们开展了一项小规模测量研究，以了解 CoinAlgs 的受欢迎程度。图 3展示了我们的结果快照，表明与传统金融一样，大量资金也投入了 Web3 CoinAlg。

AI DAO	峰值市值	峰值管理资产	隐私	主要持仓
AI16Z / Eliza	$2.7B	$22.9M	私有	ai16z, degenai, FXN
Aiccelerate DAO	$274.9M	17.1 万美元	私有	SOL、AICC、Ropirito、bAInance
大型制药人工智能	5790 万美元	69.48 万美元	私有	BIO、SOL、PYTHIA、OMIPAL
人工智能 XBT	47 亿美元	7.554 亿美元	私有	VIRTUAL ，aixbt ，ETH, CLANKER ，USDC, LUNA
Vader AI	1.426 亿美元	2.74 万美元	私有	VIRTUAL ，USDT, WETH
Axelrod	2260 万美元	2200 美元	私有	以太币、VIRTUAL、AXR、doginme
Terminal of Truths	13亿美元	2680万美元	私有	Fartcoin、GOAT、Shegen、Mana、Zerebro

图3：人工智能驱动的去中心化自治组织概览，展示其峰值市值（MC）、峰值管理资产（AUM）、是否为私有或透明，以及其主要资产持仓。

模式 #2：隐私。大多数 CoinAlgs 使用专有算法，且不对外公开。如图2所示，所有现有的机器人顾问都是私有的，因为用户通过在线界面与它们交互。同样，如图3 所示，人工智能驱动的去中心化自治组织也大多是私有的。daos.fun 等平台会提供代币估值、投资组合构成和交易历史数据，但不会公开其底层决策逻辑。

模式#3：DeFi 中的高波动投资。去中心化金融 CoinAlg 投资组合严重偏向投机性资产，并青睐新兴代币。如图3所示，许多人工智能驱动的去中心化自治组织在其主要投资组合持仓中至少包含一种“迷因币”，因此有必要研究 CoinAlg 中的稳定资产和高波动资产。

传统金融与 DeFi 中的 CoinAlgs尽管 CoinAlgs 同时存在于传统金融和 DeFi 中，但本文聚焦于后者，原因有二。首先，去中心化金融 CoinAlgs 适合进行更清晰的建模——区块链提供了自然的抽象，其复杂度远低于传统市场的抽象，也更适合开展形式化研究。其次，DeFi 的开放性质为开展 CoinAlg 的实证研究提供了天然的测试平台（例如见第6节），因为更好的数据可得性支持更贴近现实的实验和模拟。

不过，我们强调，我们的结果揭示了 CoinAlgs 中一种根本性的矛盾，这一矛盾具有广泛适用性，包括传统金融。在传统金融中，CoinAlg 困境通过法律规定的投资者保护来应对；这些保护措施使 CoinAlgs 能够保持不透明，从而保障利润。非托管 CoinAlgs 引发了严重担忧，因为它们缺乏这些保护措施。

3 CoinAlgs 的形式化模型

为从理论上研究 CoinAlg 困境，我们首先针对更广泛的 CoinAlgs 提出一个形式化模型。我们引入一个区块链系统的一般模型（借鉴了先前的研究成果 [1, 13]），随后在该模型的背景下定义 CoinAlgs，包括对 CoinAlgs 隐私的形式化定义。

系统抽象。我们用 状态 表示系统（区块链）的状态，记为 $\mathcal{S}$，并用 $\mathbb{S}$ 表示所有可能状态的集合。有效的状态转换 由函数 $\mathsf{transition}:\mathbb{S}\times\mathbb{S}\to\{0,1\}$ 定义；当且仅当第二个输入状态可从第一个输入状态到达时，该函数返回 1。我们将状态的有限序列 $\mathcal{T}=(\mathcal{S}1,\ldots\ldots,\mathcal{S}s)$ 称为状态转录，并称当对于每个 $i\in[q)$ 都满足 $\mathsf{transition}(\mathcal{S}_i,\ \mathcal{S}_{i+1})=1$ 时，该状态转录有效。

系统由一组参与者 $\mathrm{P}=\{\mathrm{P1},\ldots,\mathrm{P}_p\}$ 组成，他们在一组可交易资产 $\mathbb{A}=\{\mathrm{A1},\ldots,\mathrm{A}_a\}$ 上持有头寸。因此，特定的系统状态由所有参与者在所有资产上的余额定义。[1] 我们定义函数 $\mathsf{balance}:\mathrm{P}\times\mathbb{S}\times\mathbb{A}\to\mathbb{N}$,[2]，它将参与者映射到其在给定状态下持有的资产数量。我们用以下形式的元组表示交易：$\tau\in\mathbb{T}:=\mathrm{P}\times\mathbb{S}\times\mathbb{A}^2\times\mathbb{N}$；例如，$(\mathrm{P},\mathrm{S},\mathrm{A}_1,\mathrm{A}_2,\mathrm{N})$ 表示参与者 P 在状态 S 下将 N 个单位的 A_1 兑换为 A_2。交易 τ 所导致的状态转换由函数 $\mathsf{action}:\mathbb{T}\times\mathbb{S}\to\mathbb{S}\cup\{\perp\}$ 表示；若交易在当前状态下无效（例如，参与者持有的待售资产数量不足），该函数将返回 $\perp$。交易中获得的单位数量取决于当前资产价格；我们通过函数 $\mathsf{price}:\mathbb{A}\times\mathbb{S}\to\mathbb{R}_{>0}$ 对此建模，该函数返回以系统主要资产计价的单个资产单位成本（例如，以太坊中的以太币）。

状态转换以固定间隔触发（例如，以太坊约每 12 秒一次），此时会执行自上次状态转换以来累积的所有交易中的一部分。给定当前状态$\mathcal{S}_0$ 和一个区块，即由 $B=(\tau_1,\ldots,\tau_q)$表示的有序交易序列，最终系统状态由 $\mathsf{action}(\tau_q,\mathcal{S}_{q-1})$确定，其中$\mathcal{S}_i=\mathsf{action}(\tau_i,\mathcal{S}_{i-1})$；我们将记号扩展，并用简写$\mathsf{action}(B,\mathcal{S})$表示这一过程。请注意，如果任何底层action 调用返回 $\perp$，此过程就会返回 $\perp$，此时我们称该区块无效。区块中包含哪些交易以及它们的顺序，由特定于系统的协议决定。

CoinAlgs 模型。在我们的系统抽象中，CoinAlg$\mathcal{C}$是系统中的一个参与者，会根据

[1]从技术上讲，系统状态也由与参与者相关联的存储数据定义，但这与我们的结果无关，因此为简洁起见将其省略。[2]我们遵循如下惯例：$0\in\mathbb{N}$ 贯穿全文。

投资策略 ALG 进行交易：$\mathbb{S} \to \mathbb{T} \cup \{\perp\}$。它是一种算法，对于给定的输入状态，会返回一笔待执行的交易（如果不会进行任何交易，则返回 $\perp$）。我们用 $\mathbb{F}$ 表示所有可能的投资策略函数的集合。关键的是，$\mathcal{C}$ 的行为完全由ALG决定，即 $\mathcal{C}$ 在状态 $\mathcal{S}$ 下总会执行 $\mathsf{ALG}(\mathcal{S})$ 输出的交易。

交易算法 ALG 可以是确定性的，也可以是随机化的。为描述这一点，我们将ALG 建模为由两个逻辑子程序组成：(1) $\mathsf{ALG}_{\mathsf{dist}}(\mathcal{S})$，它（以确定性方式）输出一个关于 $\mathcal{S}$ 的概率分布 π，分布范围为 $\mathbb{T}$；随后，(2) $\mathsf{ALG}^{\mathsf{sample}}(\pi_{\mathcal{S}})$ 从概率分布 $\pi_{\mathcal{S}}$ 中抽样，得到一笔特定交易 τ。例如，确定性 CoinAlg 指的是 $\mathsf{ALG}_{\mathsf{dist}}(\mathcal{S})$ 总是输出退化分布的 CoinAlg。

CoinAlg 的隐私。现在，我们正式定义 CoinAlgs 的隐私，这是严谨分析 CoinAlg 困境的必要前提。

概括而言，我们根据参与者预测 $\mathcal{C}$ 的能力 $\mathcal{C}$ 的交易，来刻画 $\mathcal{C}$ 的隐私。请注意，我们的隐私概念是从参与者的能力（即预测交易）出发，而非依据 $\mathcal{C}$ 的具体属性（例如 $\mathcal{C}$ 的代码是否开源）。正如后文将要展示的，预测交易是 CoinAlg 困境的核心。因此，我们直接关注隐私（缺失）带来的影响，同时抽象掉信息披露的具体机制。

例如，假设 $\mathcal{C}$ 完全透明，因此参与者可以准确预测交易（的分布）。这可能是因为ALG 的规范是开源的（即参与者可以白盒访问ALG），也可能是因为参与者只有黑盒访问权限，但可以不受限制地查询ALG。对于这两种情况，关键在于参与者能够准确预测交易（正如我们在第5节所示，这会削弱 $\mathcal{C}$ 的利润），而非他们如何获取这些信息。

为了对这一点进行形式化建模，我们引入公共函数的概念，它将ALG “压平”，以刻画向公众披露的未来交易特性。（不失一般性，目前我们假设所有参与者对ALG的了解相同。[3]）具体而言，我们为 $\mathcal{C}$ 关联一个函数 $\mathcal{L}\colon \mathbb{F} \times \mathbb{S} \to \{0,1\}^* \cup \{\perp\}$，它定义了在状态 $\mathcal{S}$ 之前可以预测的关于 $\mathsf{ALG}(\mathcal{S})$ 的信息。以下是公共函数的几个示例：

(1) 完全隐私：投资策略的任何信息都不会被披露：

$$\mathcal{L}_{\mathsf{priv}}(\mathsf{ALG}, \mathcal{S}) = \perp.$$

(2) 资产透明度：仅披露交易涉及的资产：

$$\mathcal{L}_{\mathsf{asset}}(\mathsf{ALG}, \mathcal{S}) = \Big\{ A_1, A_2 \;\Big|\; (\mathcal{C}, \mathcal{S}, A_1, A_2, N) \leftarrow\!\$\, \mathsf{ALG}(\mathcal{S}) \Big\}.$$

请注意，$\mathcal{L}_{\mathsf{asset}}$ 返回的是一个无序资产集合，因此不会透露哪项资产正在出售、哪项资产正在购买。

(3) 方向透明性：交易涉及的方向和资产都会被披露：

$$\mathcal{L}_{\mathsf{dir}}(\mathsf{ALG}, \mathcal{S}) = \Big(A_1, A_2 \;\Big|\; (\mathcal{C}, \mathcal{S}, A_1, A_2, N) \leftarrow\!\$\, \mathsf{ALG}(\mathcal{S}) \Big).$$

与 $\mathcal{L}_{\mathsf{asset}}$ 不同，$\mathcal{L}_{\mathsf{dir}}$ 返回一个有序资产元组，其中第一个元素是被购买的资产，第二个元素是被出售的资产，从而揭示交易的方向。

(4) 完全透明性：整个投资策略都会被披露：

$$\mathcal{L}_{\mathsf{transp}}(\mathsf{ALG}, \mathcal{S}) = \mathsf{ALG}(\mathcal{S}).$$

公共函数是理论抽象，用于建模 CoinAlgs 向其他系统参与者揭示的行为。例如，如前所述，$\mathcal{L}_{\mathsf{transp}}$ 可能源自对ALG的白盒或黑盒访问，也可能来自更为迂回的情形，例如 $\mathcal{C}$ 以往的交易泄露了其未来将进行的交易。另请注意，取决于ALG 是确定性的还是随机化的，$\mathcal{L}$ 也具有相应的性质。

有了公开函数，我们现在可以将 $\mathcal{C}$ 的隐私定义为 $\mathcal{L}(\mathsf{ALG}\ \cdot)$ 与ALG本身之间的“距离”：距离为 0 意味着 $\mathcal{C}$ 完全透明，而距离达到最大值则意味着 $\mathcal{C}$ 完全私有。接下来，我们将具体阐述这一定义。

首先请注意，对于状态 $\mathcal{S}$，$\mathcal{L}(\mathsf{ALG}, \mathcal{S})$ 会诱导出一个概率分布 $\pi_{\mathcal{LS}}$，该分布表示参与者对于 $\mathcal{C}$ 在给定状态下可能进行何种交易的判断。例如，在完全不透明的情况下，$\pi_{\mathcal{LS}}$ 在 $\mathcal{C}$ 根据当前市场条件和 $\mathcal{C}$ 的状态（所持资产等）能够进行的所有可能交易中均匀分布；在资产透明的情况下，$\pi_{\mathcal{LS}}$ 仅在涉及所揭示的两种资产的交易中均匀分布；以此类推。基于此，我们可以将 $\mathcal{C}$ 的隐私定义为分布 $\pi_{\mathcal{LS}}$ 与 $\pi_{\mathcal{S}} \leftarrow \mathsf{ALG}_{\mathsf{dist}}(\mathcal{S})$（回想一下，这是 $\mathcal{C}$ 交易的“真实”分布）之间的统计距离 。更形式化地说：

定义 1（CoinAlg 的隐私）。设 $\mathcal{C}$ 为一个 CoinAlg，其关联的公开函数为 $\mathcal{L}$，$\pi_{\mathcal{LS}}$ 为由 $\mathcal{L}$ 对状态 $\mathcal{S}$ 诱导的 $\mathbb{T}$ 上的概率分布，且 $\pi_{\mathcal{S}} \leftarrow \mathsf{ALG}_{\mathsf{dist}}(\mathcal{S})$ 为状态 $\mathcal{S}$ 下 $\mathcal{C}$ 在 $\mathbb{T}$ 上的实际分布。我们称 $\tilde{\mathcal{C}}$ 是 ε-私有 的，如果对于任意有限状态子集 $\tilde{\mathbb{S}} \subseteq \mathbb{S}$，都有

$$\frac{1}{|\tilde{\mathbb{S}}|} \sum_{\mathcal{S} \in \tilde{\mathbb{S}}} \delta(\pi^{\mathcal{L}}_{\mathcal{S}}, \pi_{\mathcal{S}}) \geq \varepsilon$$

其中，δ 表示这对输入随机变量之间的全变差距离。

[3] 例如，我们可以将参与者各自知识的并集定义为ALG 的全局知识。

换句话说，公开信息与真实信息之间的差异越大，$\mathcal{C}$ 的隐私性就越强。需要指出的是，虽然我们使用全变差距离，但我们的隐私定义也适用于其他标准统计距离度量。

4 PrivateCoinAlgs：公平性威胁

借助我们为 CoinAlgs 建立的形式化模型，现在开始分析 CoinAlg 困境。

如第 2 节所述，实践中的 CoinAlgs 往往是私有的。这合情合理，因为将投资策略保密以保护利润是很自然的做法。然而，本节将说明，私有 CoinAlgs 会给其投资者带来经济风险。直观而言，私有 CoinAlgs 会导致信息不对称：特权实体（例如 CoinAlg 的创建者）可能完全掌握该 CoinAlg 的投资策略，并利用这一点从该 CoinAlg 的交易中提取价值。

> 关键洞见 1。私有 CoinAlg $\mathcal{C}$（其投资策略未公开披露）的投资者容易受到掌握 $\mathcal{C}$ 策略内幕知识的对手 $\mathcal{A}$ 的不公平价值提取。

我们分析的一个核心要点，是将 CoinAlgs 中的经济（不）公平性界定为一种价值提取形式，类似于去中心化金融其他领域中的最大可提取价值 [1, 8, 27, 36]，并由此自然引发经济风险。

4.1 CoinAlgs 的经济公平性定义

我们首先介绍 CoinAlgs 的经济（不）公平性定义。结合第 3 节中的隐私定义，该定义将使我们能够证明隐私与经济风险息息相关。

如第 3 节所述，参与者对 $\mathcal{C}$ 将执行的交易只有部分了解，我们通过隐私函数 $\mathcal{L}$ 对此建模。为了定义经济公平性，我们现在假设存在某个享有特权的对抗性参与者 $\mathcal{A}$，它对 $\mathcal{C}$ 的交易拥有额外的了解（形式上，拥有一个相较于 π 比 $\pi_{\mathcal{L}}$ 更接近的交易分布）。例如，$\mathcal{C}$ 的创建者就可能属于此类参与者。我们假设 $\mathcal{C}$ 对 $\mathcal{A}$ 完全透明，因此 $\mathcal{A}$ 能够提前预测交易分布；$\mathcal{A}$ 的不确定性唯一可能来源于 $\mathcal{C}$ 的随机性（例如，在可信执行环境中采样）。

经济公平性定义。我们的目标是定义 $\mathcal{C}$ 怎样才算“公平”，也就是说，没有任何参与者能够从 $\mathcal{C}$ 的交易中获取不成比例的利益。为建立直观认识，我们先从一个初步定义入手，它展示了稳健的经济公平性概念中的重要要素。

一种初步的说法是，$\mathcal{C}$ 的交易无法被恶意操纵。因此，例如，$\mathcal{C}$ 无法抽走

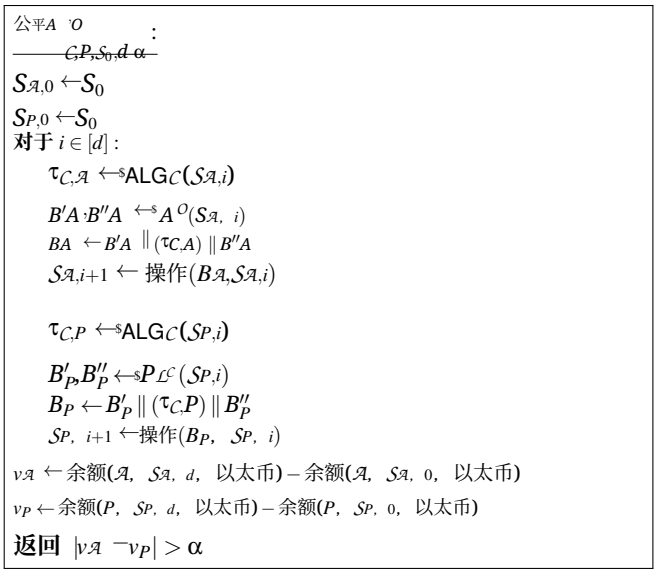

图4：CoinAlgs 经济公平性定义的博弈。

资金转入对抗性资产，也无法被操纵去进行有利于特定参与者的交易。尽管这是一个合理的威胁模型，但该定义只涵盖从 $\mathcal{C}$ 的交易中直接获利的情形，并未考虑参与者的外部能力和知识，而这些可能使他们通过内幕交易间接获利。例如，如果 $\mathcal{A}$ 知道 $\mathcal{C}$ 会在特定时间进行某笔交易，$\mathcal{A}$ 就可以对该交易发动三明治攻击，从 $\mathcal{C}$ 那里窃取利润。相反，我们希望定义能够涵盖从 $\mathcal{C}$ 的交易中可获取的总价值（直接和间接）。即使 $\mathcal{C}$ 仍然有利可图，对手也不应能够利用特权信息获取超额收益。这引出了我们定义的核心理念：将其表述为价值提取问题。

形式化地，我们通过安全博弈 FAIR 对经济公平性进行建模，该博弈定义见图4。该博弈以对手 $\mathcal{A}$ 和预言机 O 为参数，后者是一个预测预言机，用于模拟 $\mathcal{A}$ 可能拥有的辅助信息。该博弈还以参与者 P 为参数，参与者可访问 $\mathcal{C}$ 的公开函数 $\mathcal{L}$。博弈过程如下：对于从 1 到 d 的每个纪元（这是博弈的一个参数），从某个初始状态 S_0（也是博弈的一个参数）开始，$\mathcal{C}$ 会生成一笔交易，而 $\mathcal{A}$（借助 O）会生成两组交易。随后，这些交易一同执行——$\mathcal{C}$ 的交易夹在 $\mathcal{A}$ 的两组交易之间——从而使区块链转移到下一个状态。同样地，P（借助 $\mathcal{L}$）也会生成两组交易，这两组交易与 $\mathcal{C}$ 的一笔交易交错执行，以生成下一个状态。注意，在 S_0，之后，$\mathcal{A}$ 和 P 分别在独立的状态转录上进行操作。最后，如果 $\mathcal{A}$ 的余额比 P 的余额增加超过某个阈值 α，则博弈输出 1。

我们定义对手 $\mathcal{A}$ 相对于 CoinAlg $\mathcal{C}$ 的优势

（其关联的公开函数为$\mathcal{L}$），针对预言机$\mathcal{O}$、参与者P、初始状态$\mathcal{S}_0$, 持续时间d 和阈值 α，定义如下：

$$\mathrm{Adv}^{\mathrm{fair}}_{\mathcal{C},\mathcal{O},P,\mathcal{S}_0,d,\alpha}(\mathcal{A}) = \left|\Pr\left[\mathrm{FAIR}^{\mathcal{A},\mathcal{O}}_{\mathcal{C},P,\mathcal{S}_0,d,\alpha} = 1\right]\right|.$$

现在我们给出经济公平性的定义：

定义2 (CoinAlgs 的经济公平性)。 若对于任意对手$\mathcal{A}$、状态 $\mathcal{S} \in \mathbb{S}$ 和整数 $d \in \mathbb{N}$，存在一个高效算法 P，使得 CoinAlg$\mathcal{C}$ 关于预言机$\mathcal{O}$ 是 (α,t)-公平的，则称之为如此。

$$\mathrm{Adv}^{\mathrm{fair}}_{\mathcal{C},\mathcal{O},P,\mathcal{S}_0,d,\alpha}(\mathcal{A}) \le t.$$

换言之，我们的公平性定义刻画了$\mathcal{A}$的预测预言机 $\mathcal{O}$ 的能力，具体而言，它能让$\mathcal{A}$提取多少额外利润。

如本节前文所述，我们最关注的情况是 $\mathcal{O}$ 能够完美预测 ALG，这是定义 2 的一个特殊情形。对于给定的起始状态和参数集，我们将$\mathcal{A}$ 使用完美预测预言机可提取的价值称为完全知情可提取价值。

定义2 以参数化方式对$\mathcal{C}$的（不）公平性进行建模，因为它是根据价值提取差异的特定阈值和对手成功概率来表述的。因此，更准确地说，它捕捉的是$\mathcal{C}$整体公平性的一个快照。因此，我们的定义旨在灵活使用，可针对不同的 (α,t) 多次应用。概念上，这将形成一条公平性曲线，从更广泛的角度刻画$\mathcal{C}$的公平性。

4.2 CoinAlgs 中的公平性与隐私

现在我们继续讨论本节的主要结果，论证 CoinAlgs 中的隐私与经济公平性之间存在张力。具体而言，我们证明，有意义的信息预言机（即允许对手提取额外价值的预言机）一旦存在，就意味着 $\mathcal{C}$ 必须是私有的，因此隐私是不公平性的前提。

定理 1。 设$\mathcal{C}$ 是一个 CoinAlg，针对完美预测预言机$\mathcal{O}$满足 (α,t)-不公平性，其中 $\alpha,t > 0$。那么，存在某个 $\varepsilon > 0$，使得 $\mathcal{C}$是 ε-私有的。

我们将该定理的证明推迟至附录A。直观来说，如果$\mathcal{A}$能够利用$\mathcal{O}$，比任何使用$\mathcal{L}$的P 提取更多价值，那么$\mathcal{O}$ 和$\mathcal{L}$ 之间必定存在信息不对称，从而导致价值差异；根据定义，这意味着$\mathcal{C}$ 是私有的。更形式化地说，如果$\mathcal{A}^{\mathcal{O}}$ 相较于$\mathcal{L}P$提取了更多价值，那么试图区分$\mathcal{O}$和$\mathcal{L}$的区分器*distinguisher* Dist （这是定义1的一种等价表述）可以将$\mathcal{A}$和P 作为子程序运行，并利用提取价值的差异来判断它正在与哪个预言机交互。

请注意，定理1 是一个方向性结果。未来值得探讨的一个有趣问题是研究细粒度变体，并量化对于定义2 的特定参数值，隐私会在多大程度上降低。

隐私是否意味着不公平性？ 接下来一个自然的问题是，定理 1 的逆命题是否也成立，即隐私是否总会导致经济不公平性。然而，事实并非总是如此：额外的信息并不总是带来额外的价值提取，因为从经济视角来看，信息并非总是有用的。因此，定理 1 的逆命题一般而言并不成立。

然而，如果情况是 $\mathcal{C}$ 以特权信息具有经济价值的方式保持私有，那么这将直接导致经济不公平性。这一推论是有意为之，因为我们的 FAIR 博弈正是为了精确模拟信息预言机的金融效用而定义的。因此，在我们的 FAIR 博弈中，可以提取出更多价值的预言机（按定义）会导致更大的价值差距。

现在让我们将这一想法形式化。我们将信息预言机的金融效用 建模为其能够实现的完全信息可提取价值的百分比 （如前文所定义）。根据定义，完美预测预言机 $\mathcal{O}$ 的金融效用为 100%。于是我们得到以下简单结果。

定理 2. 设 $\mathcal{C}$ 是一个 CoinAlg，其关联的公共函数 $\mathcal{L}$ 的金融效用为u。假设 $\mathcal{C}$ 相对于某个 $\varepsilon > 0$ 的完美预测预言机 $\mathcal{O}$ 是 ε-私有的。那么，存在某个 $\alpha,t > 0$，使得 $\mathcal{C}$ 相对于完美预测预言机$\mathcal{O}$是 (α,t)-不公平的。

如前所述，这一结果是由构造保证的：如果$\mathcal{L}$的金融效用为u，那么$P^{\mathcal{L}}$ 在 FAIR 中最多只能提取uT 的价值，其中T 是ALG的完全信息可提取价值。另一方面，$\mathcal{A}^{\mathcal{O}}$ 则能够在 FAIR 中提取T 的价值。

我们指出，信息的具体价值高度依赖于情境。因此，要得出更普遍、更细致的界限，衡量（有用的）信息会在多大程度上加剧经济不公平，很可能并不可行。

5 透明硬币算法：对盈利能力的威胁

鉴于我们在上一节关于私有 CoinAlgs 中经济不公平风险的研究结果，CoinAlgs 采用透明设计似乎是更好的选择。毕竟，透明性是确保系统可信度的一项关键属性，适用于从开源软件到智能合约等各类系统。

然而，在本节中我们将表明，CoinAlg 中的完全透明性——即公开知晓ALG——是以牺牲盈利能力为代价的。原因很直观：透明性意味着向市场公开发出交易信号。套利者可以利用这些提前获得的信息获利，代价则由

CoinAlg 承担。本节的工作总体上为以下观点提供了证据：

> 关键洞见 2。 一个策略 为ALG 的 CoinAlg $\mathcal{C}$ 是透明的——套利对手$\mathcal{A}$知晓该策略——因此会产生成本为透明性成本：其利润低于等价的私有 CoinAlg $\mathcal{C}^*$。

我们所说的等价私有 CoinAlg $\mathcal{C}^*$，是指$\mathcal{C}^*$执行与 ALG 相同的策略，但$\mathcal{C}$的ALG 对$\mathcal{A}$ 未知，因而不会受到抢先交易/三明治攻击的影响。$\mathcal{C}$的透明性成本 是指其利润低于$\mathcal{C}^*$的差额。（详情如下。）

即使透明性成本很小，也可能使$\mathcal{C}$处于竞争劣势。这种劣势可能导致$\mathcal{C}$的市场失灵，因为投资者会转向私有替代方案所带来的更高回报。因此，关键洞见 2 极大地推动了人们对私有 CoinAlgs 进行利润最大化的兴趣。结合我们在第 4 节关于私有 CoinAlgs 风险的研究结果（关键洞见 1），显然，无论如何设计 CoinAlgs，都存在重大风险。

5.1 对抗策略

对手 $\mathcal{A}$ 会对透明的 CoinAlg $\mathcal{C}$构成两种截然不同的策略威胁：

威胁 1（策略窃取）：如果$\mathcal{C}$即将执行交易 τ_1，$\mathcal{A}$ 可以复制并抢先交易，在 τ_1之前执行一笔等效交易 τ_0。这样一来，$\mathcal{A}$ 窃取$\mathcal{C}$的投资策略。

示例：预测 TOK 的价格将从现在的 10 涨至明天的 15，$\mathcal{C}$ 准备在交易 τ_1中买入 500 TOK。$\mathcal{A}$ 利用$\mathcal{C}$的透明性，在 τ_1之前执行 $\tau_0 = \tau_1$ ——如果$\mathcal{C}$的预测正确，便可获利；而$\mathcal{A}$的利润可能会因此受损，因为 τ_0 会抬高$\mathcal{C}$买入 TOK 的价格。

威胁 2（三明治攻击）：$\mathcal{A}$ 可以利用$\mathcal{C}$的透明性，对$\mathcal{C}$的交易 τ_1实施三明治攻击——执行配对的抢先交易和尾随交易。这样一来，$\mathcal{A}$ 便能利用 τ_1 引发的价格变动信息提前获利，同时削减$\mathcal{C}$的利润。

示例：给定与上文相同的 τ_1 （买入 500 TOK），$\mathcal{A}$利用$\mathcal{C}$的透明性执行 τ_0，在 τ_1 之前买入 5 TOK，并在 τ_1 之后于 τ_2 卖出 5 TOK。$\mathcal{A}$ 因此获利，因为τ_1 会在$\mathcal{A}$ 卖出之前推高 TOK 的价格，而 τ_0 会推高$\mathcal{C}$ 在 τ_1 中支付的 TOK 价格，从而降低$\mathcal{C}$的利润。

对手很容易实施策略窃取，因为这只涉及复制。不过，这会给$\mathcal{A}$带来难以量化的风险，因为$\mathcal{C}$的预测可能不准确，因此其策略在任何给定时间段内都可能盈利，也可能无法盈利。我们在第6 节进行实验，以评估策略窃取的影响。

本节余下部分将聚焦于威胁 2。与策略窃取相比，三明治攻击策略更适合进行解析研究。它们可以是无风险的，也就是说，它们并不依赖$\mathcal{C}$策略的可靠性，而是能为$\mathcal{A}$带来即时利润。正因如此，三明治攻击如今已成为获取最大可提取价值的一种广泛策略 [1, 8, 35, 36]。

三明治攻击只会影响透明 CoinAlgs 吗？ 三明治攻击——以及更广泛的最大可提取价值提取——会降低任何暴露给套利者的交易的利润，无论这些交易位于公共还是私有内存池中。因此，私有 CoinAlg也可能 容易受到攻击。不过，私有 CoinAlg可以利用三明治攻击防护措施，例如 Flashbots Protect [10]、Cowswap [6], 等服务，或与验证者达成侧面协议。如果$\mathcal{C}$ 是透明的，这些防护措施就无法 奏效，因为执行的交易会泄露给$\mathcal{A}$。

5.2 模型细节

我们首先为针对透明CoinAlg的三明治攻击引入一个形式化模型，并在第 3 节的形式化框架基础上加以扩展。该模型将构成后续结果的基础。

市场模型。我们考虑一个简单市场，其中两种资产在单个恒定乘积自动做市商内进行交易。为便于说明，我们将这两种资产记为 TOK 和美元（例如以太币和一种通用的美元稳定币）。我们用 t 表示该自动做市商的逻辑时间——即其已执行交易次数的计数器。时间步 t 执行之前（或时间步 $t-1$ 之后）的自动做市商状态，表示为交易对余额 $\mathrm{bal}^{\mathsf{AMM}}\mathrm{t} = [\mathrm{p}_t\mathrm{r}_t\ \mathrm{USD}, \mathrm{r}_t\ \mathrm{TOK}]$。在恒定乘积自动做市商中，这对应于 TOK-USD 市场价格 $(\mathrm{p}_t\mathrm{r}_t)/\mathrm{r}_t = \mathrm{p}_t$。我们的时间模型与第 4 节中的模型一致，但假设交易仅涉及该自动做市商上的 TOK-USD 交易对。下文中，我们使用“交易”一词指代在应用层执行交易的完整数字签名载荷 τ，也就是自动做市商操作。

交易分为两种：用美元买入 TOK，或卖出 TOK 换取美元。为简化第 4 节中的记号，我们用 买入(x) 表示买入 x TOK，用买入$(-x)$表示卖出xTOK。我们假设自动做市商或智能合约支持根据自动做市商中的 TOK 价格进行条件交易执行，即支持限价单。我们将限价为$p_{\lim}$ 的限价单记为 买入$(x)\downarrow_{p_{\lim}}$，表示只要 TOK 价格不高于$p_{\lim}$，交易就会执行。[4]

[4]自动做市商交易通常指定的不是价格上限，而是滑点限制。价格上限简化了我们的分析，也可以通过智能合约轻松执行，正如第6节中的实验所示。

C 的策略模型。我们关注的是具有跑赢市场表现的 C 的算法 ALG，这也是投资者对 CoinAlg 感兴趣的原因。我们考虑一个简单模型：内部预言机 $\hat{O}[S_i]$ 会针对某个未来时刻 $\hat{t}$ 发布对 $\hat{}$ TOK-USD 价格 $\hat{p}$ 的预测。[5] 我们将 ALG 中使用该预言机称为 ALGO。（为便于阅读，在适当情况下，我们写作 ALGO(S_i)，甚至直接写作 ALGO，而不写作 ALG $\hat{O}[S_i]$(S_i)。）正是 $\hat{O}$（预测 $\hat{p}$ 的来源）体现了 C 跑赢市场的策略——也就是它的“独门秘诀”。

为简化起见，我们在此假设 ALG 在后文讨论放宽这一假设的可能性。此外，不失一般性，我们假设 $\hat{p}$ 高于当前以美元计价的 TOK 价格。（如果 $\hat{p}$ 较低，那么 $\mathcal{C}$ 可以将美元视为未来价格较高的目标资产。）

$\mathcal{C}$ 的策略很简单，是单向的。其目标是通过重复以下过程，使该投资组合的美元价值最大化。给定当前价格 p 以及对未来时间 $\hat{t}$ 时 TOK 价格的预测 $\hat{p} > p$：

- 执行买入 $(u)_{\downarrow p_{\text{lim}}}$ 数量的 TOK，其中 u 会将自动做市商价格推高至预测价格 $\hat{p}$。
- 在时间 $\hat{t}$ 卖出 u 单位的 TOK。

展望未来，我们在第 6 节中的实证分析将考察更加复杂的交易策略。

为简化起见，我们假设出售 TOK 能够完美实现 $\mathcal{C}$ 预测的市场价格，即对于 TOK-USD 交易，我们不考虑滑点、抢先交易等因素。

透明性模型。 如第3节所定义，如果每个参与者（包括 $\mathcal{A}$）都能访问公开函数 $\mathcal{L}_{\text{transp}}(\text{ALG}_{\mathcal{O}}) = \text{ALG}_{\mathcal{O}}(\cdot)$，则 $\mathcal{C}$ 是透明的。请回想，这一定义建模了对 $\text{ALG}_{\mathcal{O}}(\cdot)$ 的无限查询访问，但我们不对这种访问在实践中如何实现作任何假设；$\mathcal{L}_{\text{transp}}$ 是对任何至少提供 $\text{ALG}_{\mathcal{O}}(\cdot)$ 黑盒访问的环境的抽象。这可以通过访问 $\mathcal{A}$ 能够查询的某个接口、$\mathcal{C}$ 的开源实现等方式实现。

一个推论是——鉴于我们假设 $\mathcal{C}$ 具有确定性——$\mathcal{A}$ 可以在交易执行之前预测 $\text{ALG}_{\mathcal{O}}$ 的交易。“之前”究竟意味着什么，取决于我们在博弈定义和实验中考虑的具体情景。例如，后文将介绍的博弈Sdwch假设 $\mathcal{A}$ 能够及时获知 τ_1，从而成功对 $\mathcal{C}$ 发动三明治攻击。实际上，对于大多数区块链，这意味着 $\mathcal{A}$ 可以在 $\mathcal{C}$ 获知 τ_1 的同时甚至稍晚一些才获知该信息，但仍能将自己的交易安排在区块中更早执行，以完成交易 τ_0；如今的 MEV 提取正是如此。此能力直接源于透明性，因为 $\mathcal{A}$ 只需与 $\mathcal{C}$ 并行执行 $\text{ALG}_{\mathcal{O}}$ 即可。

[5] $\hat{O}$ 可能会引用链下数据。由于我们假设 $\text{ALG}_{\hat{O}}$ 具有透明性，$\hat{O}[S_i]$ 的输入都是公开可用的，并按 i 进行规范索引，因此任何参与者——尤其是 $\mathcal{A}$——都可以计算 $\hat{O}[S_i]$。

对手 $\mathcal{A}$ 的模型。 我们将 $\mathcal{A}$ 建模为单体实体，即单一实体。实际上，$\mathcal{A}$ 可以代表一群合作或独立的套利者。我们关注的是 $\mathcal{C}$ 如何损失利润，因此这种建模足以支持我们的结论；个人如何获利则是另一个无关的问题。对于 $\mathcal{A}$ 的一对三明治交易（τ_0，τ_2），我们的单体模型给出了 $\mathcal{A}$ 利润的上界，因为套利者之间的竞争可能会降低 $\mathcal{A}$ 的利润（例如，通过 MEV 拍卖成本）。然而，我们的模型对 $\mathcal{C}$ 因三明治攻击而遭受的损失给出了紧确界，因为套利者会集体最大化从 $\mathcal{C}$ 提取的价值，将自动做市商推至 $\mathcal{C}$ 的限价 p_{lim}，其中 τ_0。[6] 我们将时间步 t 执行前 $\mathcal{A}$ 的投资组合记为 $\text{bal}_{\mathcal{A}t} = [c_t \text{ 美元}, d_t \text{ TOK}]$。

为了衡量三明治攻击对（透明）$\mathcal{C}$ 利润的影响，我们将其与这样一种情形进行比较：私有CoinAlg $\mathcal{C}^*$ 也执行相同的策略ALG；也就是说，$\mathcal{A}$ 转而只能访问公开函数 $\mathcal{L}_{\text{priv}}(\text{ALG}\ \cdot) = \perp$，因此 $\mathcal{A}$ 无法提前获知 $\mathcal{C}$ 的交易。因此，我们通过假设 $\mathcal{A}$ 不会发动三明治攻击来建模这一私有情形。（这里还隐含着一个假设：$\mathcal{C}^*$ 提交交易时不会将其暴露在内存池中——在实践中，可以使用例如 Flashbots Protect [10]这样的服务做到这一点。）更正式地说：

定义 3 （透明性成本）**.** 设 $\Pi_{\mathcal{C}}$ 表示某个给定博弈在均衡时透明CoinAlg $\mathcal{C}$ 的收益，$\Pi_{\mathcal{C}^*}$ 表示其私有版本 $\mathcal{C}^*$ 的收益。我们将 $\mathcal{C}$ 的透明性成本定义为 $\Pi_{\mathcal{C}^*} - \Pi_{\mathcal{C}}$。

也就是说，我们将不存在三明治攻击的情形作为基准，并根据不同情形下 $\mathcal{C}$ 的收益与 $\mathcal{C}^*$ 的收益偏离多少来进行比较。

区块链模型。 区块构建、围绕区块链的基础设施和市场在不同区块链及其设计演进之间存在很大差异。为保持一般性，我们抽象掉区块构建，仅假设 $\mathcal{A}$ 能够对交易进行有利排序——无论是通过使用 MEV供应链、利用 $\mathcal{C}$ 的透明性，还是其他方式。

5.3 建模说明

在进入我们的结果之前，我们先对模型作几点澄清说明。

备注 1 （交易与交易泄露）。 在第 3 节的透明性模型中，τ_1 中的交易，即自动做市商操作，会泄露给 $\mathcal{A}$。交易本身 τ_1 及其数字签名则未必会泄露。（在自然架构中，$\mathcal{C}$ 的钱包将独立于 $\text{ALG}_{\mathcal{O}}$。）因此，在 MEV 供应链中，$\mathcal{A}$ 未必

[6] 在 MEV供应链中，$\mathcal{C}$ 可能会竞价获取有利的区块头寸，以防止遭受 $\mathcal{A}$ 的三明治攻击。不过，$\mathcal{C}$ 需要对抗 $\mathcal{A}$，将出价提高至最高 $s - \varepsilon$，其中 s 是 A 通过三明治攻击获得的收益。为进行与区块链无关的分析，我们的模型排除了这种可能性；但请注意，这将导致与我们模型中所展示的相同的高透明性成本。

能够通过捆绑，以原子方式执行序列（τ_0，τ_1，τ_2）。$\mathcal{A}$ 仍然可以 以非原子方式对$\mathcal{C}$ 实施三明治攻击，这大致相当于在我们的单体模型中对$\mathcal{A}$进行原子三明治攻击。

备注 2（（弱）透明性假设）。 实际上，在$\mathcal{L}^{\text{transp}}$下，在交易执行之前预测交易是可行的，这只需对对手的能力作出较弱的假设：$\mathcal{A}$ 只需通过黑盒ˆ访问即可提前获知 τ_1，而无需逆向工程ALGo，即无需拥有白盒访问权限。

如果$\mathcal{C}$根据实时信息生成交易，$\mathcal{A}$就可以如上所述，通过与$\mathcal{C}$并行执行ALGo来发动三明治攻击。或者，如果$\mathcal{C}$进行中长期价格预测，锁定交易 τ_1 后等待有利的执行条件，则可以通过模拟未来可能的区块链状态，并在各次模拟中始终由ALGo 输出 τ_1ˆ作为交易，从而提取 τ_1。

通过完整披露ˆALGo，当然可以实现黑盒访问，但即使仅拥有（广泛的）查询访问权限，也同样可以实现。

备注 3（确定性与非确定性）。 尽管我们在此假设$\mathcal{C}$具有确定性——这是一个强假设——但我们认为，我们的结果对于许多自然形式的非确定性$\mathcal{C}$也同样具有方向性意义。原因在于，$\mathcal{C}$能够实现最大利润的交易范围是有限的。例如，一个以获取利润为目的进行交易的$\mathcal{C}$不太可能在买入TOK 和卖出TOK 之间随机选择。它也不太可能在大量资产中随机选择，或大幅改变交易规模。[7] 因此，$\mathcal{A}$ 可能无法准确预测 τ 确切结果，但很可能能够很好地近似它，而这足以利用该结果。备注2 在此意义上同样适用于非确定性的$\mathcal{C}$。我们在第6节通过实证验证了这一主张：在我们的实验中，即使对 τ的了解并不完整，A仍能够显著降低$\mathcal{C}$的利润。

在有利参数范围内随机化交易选择，可能提供一定防御手段，以抵御$\mathcal{A}$的利用——这也是我们在第7节提出启发式防护机制的依据。

5.4 基于博弈论的三明治攻击模型

为评估$\mathcal{C}$与$\mathcal{A}$之间的互动，我们将三明治攻击场景建模为斯塔克尔伯格博弈 [32]，参与者为$\mathcal{C}$ 和$\mathcal{A}$。我们将其称为Sdwch，具体如图5所示。斯塔克尔伯格博弈刻画领导者与跟随者之间的互动，二者依次行动：领导者先行动，跟随者再做出响应。因此，这类博弈为三明治攻击建模提供了自然的抽象。在我们的场景中，一个微妙之处在于，尽管$\mathcal{A}$的交易 τ_0 首先执行，$\mathcal{C}$ 仍然充当领导者，因为$\mathcal{A}$ 会根据$\mathcal{C}$的交易 τ_1来决定其三明治攻击交易。

单次斯塔克尔伯格博弈Sdwch

博弈设置：

- 初始状态：S
- 初始余额：

$$\begin{aligned} \text{余额}_0^{\text{AMM}} &= [p_0 r_0 \text{ 美元}, r_0 \text{TOK}] \\ \text{余额}_{\mathcal{C}0} &= [a_0 \text{ 美元}, b_0 \text{ TOK}] \\ \text{余额}_{\mathcal{A}0} &= [c_0 \text{ 美元}, d_0 \text{ TOK}] \end{aligned}$$

- 预测价格：$\hat{p} \leftarrow \hat{O}(S)$

游戏序列：

1. **领导者$\mathcal{C}$：** 选择交易 τ_1：

$$\tau_1 = \text{buy}(u)_{\downarrow p_{\text{lim}}} \leftarrow \text{ALG}^{\hat{O}}(S).$$

2. **跟随者$\mathcal{A}$：** 选择三明治攻击交易 τ_0, τ_2：

$$(\tau_0, \tau_2) \leftarrow \mathcal{A}^{\text{ALG}^{\hat{O}}}(S), \quad \text{where } \tau_0 = \text{buy}(x) \text{ and } \tau_2 = \text{buy}(-x).$$

自动做市商（自然）执行：

(1) τ_0，（2）τ_1 当且仅当$p_1 \le p_{\text{lim}}$，（3）τ_2.

$\mathcal{C}$、$\mathcal{C}^*$和$\mathcal{A}$的收益（单位：美元）：

- $\Pi_{\mathcal{C}}$：$u \times \hat{p} - (a_0 - a_3)$ - 费用
- $\Pi_{\mathcal{C}^*}$：$u \times \hat{p} - (a_0 - a_3)$ - 费用 对于$x = 0$（无$\mathcal{A}$）
- $\Pi_{\mathcal{A}}$：$c_3 - c_0$ - 费用

图 5：双步斯塔克尔伯格博弈 Sdwch，用于建模透明 CoinAlg $\mathcal{C}$（作为领导者）与套利者$\mathcal{A}$（作为跟随者）之间的互动。$\mathcal{C}$选择交易 τ_1，而$\mathcal{A}$对其发动三明治攻击。$\mathcal{C}$的收益 $\Pi_{\mathcal{C}}$是其边际投资组合价值：所购TOK的预测美元价值减去美元成本（对于$\mathcal{C}^*$，不包括$\mathcal{A}$发动三明治攻击的情况）。$\mathcal{A}$的收益是纯美元利润。

如前所述，为了限制$\mathcal{A}$利用$\mathcal{C}$的风险，$\mathcal{C}$可以 τ_1 提交一个带有限价p_{lim}的限价单。随后，$\mathcal{C}$ 会以不高于未来价格 $\hat{p}$的价格购买TOK。p_{lim} 的取值对我们的均衡分析很重要。然而，由于$\mathcal{C}$完全透明，$\mathcal{A}$ 事先就知道p_{lim}的取值，因此$\mathcal{A}$ 可以相应地应对$\mathcal{C}$的这一防御措施。

单阶段博弈。 在一次Sdwch重复中，$\mathcal{C}$有一种自然的策略：将 τ_1 的限价设为$p_{\text{lim}} \leftarrow p_0 + \varepsilon_{\text{min}}$，其中$p_0$ 是执行 τ_0 之前的 TOK 价格，ε_{min} 是最小价格增量。也就是说，$\mathcal{C}$ 采用保守防御策略，一旦套利者试图发动三明治攻击，就立即中止交易。此策略使$\mathcal{C}$在斯塔克尔伯格均衡中获得最优收益。（若设置$p_{\text{lim}} = 0$，则$\mathcal{A}$ 可以零成本地剥夺$\mathcal{C}$ 的收益，

[7] 将 大额交易拆分成小额交易是常见的策略，旨在 优化执行，但滑点、费用等限制了策略 。

正如稍后讨论的重复博弈设定中那样。）$\mathcal{C}$ 几乎获取了全部剩余，因为$\mathcal{A}$ 只有在价格不超过 $\varepsilon_{\min}$时才能实施三明治攻击。我们用以下简单定理来说明这一点：

定理 3 ($\mathcal{C}$ 在单次游戏Sdwch 中的透明性成本为零。)。在Sdwch的单次变体中，唯一的均衡是$\mathcal{C}$的透明性成本为零，即 $\Pi_{\mathcal{C}^*} - \Pi_{\mathcal{C}} = 0$。

这显然是因为，既然$\mathcal{A}$的交易至少会使价格上涨 $\varepsilon_{\min}$，它就无法通过对该交易发动三明治攻击来提高自身收益。

重复博弈。 然而在实践中，CoinAlgs 和抢跑者会进行长期的多区块交互，这会彻底改变均衡分析。因此，我们现在转向重复变体Sdwch。与单次情形的关键区别在于，正如重复博弈中的常见做法，参与者现在可以使用触发策略 [22] 来迫使参与者合作。这类策略会在未来回合中构成可信的惩罚威胁，从而激励参与者在当前回合中合作。

我们的核心洞见是，$\mathcal{A}$ 可以将使交易失效武器化，作为冷酷触发策略。也就是说，$\mathcal{A}$ 可以执行一笔交易 τ_0，买入x TOK，并将 TOK 的美元价格从当前价格p 推高至高于 p_{lim}（或反向操作），从而导致$\mathcal{C}$执行的交易 τ_1 失败。因此，如果$\mathcal{C}$ 不提高p_{lim} 并分享剩余，$\mathcal{A}$ 就可以威胁使$\mathcal{C}$ 的所有未来交易失效，并剥夺其未来利润，从而通过惩罚威胁迫使其合作。（为简化分析，我们重点研究冷酷触发策略，但我们的结果也适用于更细粒度的触发策略。）

触发策略下的均衡将严重取决于对手进行交易失效的成本。最坏情况下，$\mathcal{A}$ 会因套利／MEV 拍卖成本而损失其 TOK 的边际价值（等于 $x(p_{\text{lim}} - p)$），并支付区块链交易费用（本质上为常数c）以及自动做市商交易费用（交易规模的固定百分比μ，因此为μp_{lim}）。最好的情况下，$\mathcal{A}$ 只需承担这些费用。我们在下面对其进行形式化量化：

引理 1。 在Sdwch中，如果对手$\mathcal{A}$ 通过 τ_0 =买入(x)进行足够幅度的抢先交易，将自动做市商价格从p 提高到 $> p + p_{\text{lim}}$（因此 τ_1 失败），则$\mathcal{A}$的交易成本大于

$$C(p_{\text{lim}}) := p r_0 \left(\sqrt{\frac{p + p_{\text{lim}}}{p}} - 1 \right) \text{ USD.} \qquad (1)$$

我们在附录B.1中给出了该引理的证明。公式1 可视为对$\mathcal{A}$ 施加的惩罚，因为其使$\mathcal{C}$的交易失效。因此，p_{lim} 与Sdwch中的收益之间存在权衡：p_{lim} 越低，$\mathcal{C}$可能提取的利润就越高，但$\mathcal{A}$ 使该交易失效的成本也越低。

我们的目标是在冷酷触发策略下分析Sdwch，并证明$\mathcal{A}$ 能够迫使$\mathcal{C}$ 达到一种必须增加p_{lim}的均衡。然而，直接分析Sdwch 十分复杂，

最后通牒博弈 U-Sdwch

博弈序列：

1. **领导者$\mathcal{C}$：** 将剩余s 分配给自己和$\mathcal{A}$:
$$(e, s-e) \leftarrow \mathcal{C}, \text{ 其中 } 0 \le e \le s$$
2. **跟随者 $\mathcal{A}$：** 接受或拒绝剩余的分配:
$$b \leftarrow \mathcal{A}(e), \text{ 其中 } b \in \{0,1\}$$

$\mathcal{C}$、$\mathcal{C}^*$ 和$\mathcal{A}$ 的收益（单位：美元）：

- $\Pi_{\mathcal{C}}$： $bc_1(s-e)$ −费用
- $\Pi_{\mathcal{A}}$： $bc_2 e - (1-b) z\sqrt{e}$− 费用

常数 $c_1 > c_2 > 0$ 和 $z > 0$。

图6：最后通牒博弈 U-Sdwch，这是博弈 Sdwch 的一种风格化变体。CoinAlg $\mathcal{C}$ 向$\mathcal{A}$提出对可用剩余的分配方案；$\mathcal{A}$可以接受该方案，也可以拒绝，从而使双方都无法获得利润（但拒绝的代价与向其提供的份额成正比）。

由于状态变化涉及许多外生因素（超出$\mathcal{C}$ 与$\mathcal{A}$互动的范围），对Sdwch 的完整分析需要考虑这些环境因素带来的变化。因此，我们将转而分析Sdwch的一个风格化变体 中的均衡，该变体抽象掉外部因素，并将博弈简化为其关键组成部分。

5.5 风格化三明治攻击模型

其核心在于，$\mathcal{C}$ 与$\mathcal{A}$ 之间的互动涉及一笔可用的剩余（源于$\mathcal{C}$的价格预测），$\mathcal{C}$ 通过设定p_{lim}在自身与$\mathcal{A}$ 之间分配这笔剩余。随后，$\mathcal{A}$ 可以选择接受或拒绝这一分配（即是否使交易失效）。这种情形自然可以建模为最后通牒博弈 [12]。这类博弈在博弈论文献中十分常见，其中有一名提议者提出一笔金额的分配方案，另一名回应者接受或拒绝该分配。这类博弈与$\mathcal{C}$ 和$\mathcal{A}$之间的抽象互动直接对应。我们在图6中展示了风格化的最后通牒博弈U-Sdwch。U-Sdwch 十分简单：有一笔可用的剩余s，$\mathcal{C}$ 在某个截止点e处分配这笔剩余，将e 交给$\mathcal{A}$，并将剩余的$s-e$ 留给自己。随后，$\mathcal{A}$可以选择接受该分配，或烧掉全部剩余，这由其选择的比特b表示。

收益结构。 定义我们的风格化博弈时，一个挑战是确定收益结构，即一般 Sdwch博弈中的收益如何映射到这一更简单的抽象中。特别要注意的是，我们的情形不同于标准的最后通牒博弈，因为 $\mathcal{A}$ 必须 付费以拒绝该分配，因为使$\mathcal{C}$的交易无效需要承担成本（引理 1）；在最后通牒博弈中，拒绝通常会导致收益为 0，

但在我们的情形中并非如此。现在让我们分析$\mathcal{A}$接受分配时以及不接受时的收益。

如果$b = 1$，$\mathcal{C}$的交易未被取消，因此它可以从自己的未来价格预测中获益，而$\mathcal{A}$可以提取三明治攻击的即时价值。这直接对应于我们的简化模型：$\mathcal{C}$的收益为$(s-e)\hat{}p$，A的收益为ep，我们可以将其简化为<style id='c'>$c_1(s-e)$和$c_2 e$，其中$c_1 \geq c_2 > 0$为常数。

如果 $b = 0$，$\mathcal{C}$ 获得的收益为 0，而 $\mathcal{A}$ 需支付方程1中的失效成本。将方程1 映射到我们的简化博弈中，p_{lim} 对应于$p+e$，因此$\mathcal{A}$拒绝拆分的成本为

$$C(e) \coloneqq p r_0 \left(\sqrt{\frac{2p+e}{p}} - 1 \right).$$

注意到$C(e)$随着 $\sqrt{e}$缩放，因此我们可以将该表达式简化为 $\sqrt{e}$的函数：

引理 2。 对于任意$e \geq 0$，存在一个常数z 使得

$$C(e) \leq z\sqrt{e}$$

我们在附录B.2中给出了该引理的证明。我们使用这个失效成本上界，因为它使我们能够简化均衡分析，同时考虑一个事实上对$\mathcal{C}$更有利的情形。

综上，$\mathcal{C}$和$\mathcal{A}$ 在U-Sdwch中的收益为：

$$\Pi_{\mathcal{C}} \coloneqq bc_1(s-e) + (1+b)0 - \text{fees, and}$$
$$\Pi_{\mathcal{A}} \coloneqq bc_2 e + (1-b)(-z\sqrt{e}) - \text{fees.}$$

请注意，b 就像一个开关，根据$\mathcal{A}$ 是否接受剩余，来决定各方的收益。

重复博弈均衡。 现在我们分析U-Sdwch的重复博弈变体，并说明$\mathcal{A}$ 可以采用冷酷触发策略来维持一种$\mathcal{C}$ 分享剩余的均衡。然而，关键在于，$\mathcal{A}$的冷酷触发策略必须构成一种可信威胁，才能迫使对方合作：如果$\mathcal{A}$ 没有理性理由采用冷酷触发策略，$\mathcal{C}$ 就会独占全部剩余，无视这一威胁。不过，我们的情形有一个微妙之处：如前所述，冷酷触发策略会带来成本，因此$\mathcal{A}$ 实际上接受任意 比例的s都能获得更高收益，即使该比例再小也一样。因此，事先看来，$\mathcal{A}$ 采用冷酷触发策略并不理性，也就无法构成可信威胁。

为了解决这一问题，$\mathcal{A}$ 可以采用一种标准方法：在博弈开始前承诺采用冷酷触发策略。通过使用承诺装置 [17] 来约束冷酷触发策略的使用，$\mathcal{A}$ 可以促使$\mathcal{C}$ 行事时仿佛冷酷触发策略是一种可信威胁，即便在之后 $\mathcal{C}$背叛时才决定采用该策略并不理性。已有大量研究探讨如何在各种情境下对策略作出承诺 [2]；本节的结果并不取决于$\mathcal{A}$采用的具体机制。尤其是，区块链通过智能合约提供了天然的承诺装置 [11]。

现在我们得出了主要结果：$\mathcal{A}$ 可以采用冷酷触发策略，迫使$\mathcal{C}$ 分享剩余。

定理 4. 在U-Sdwch的重复变体中，存在一个$e^* > 0$，使得$\mathcal{A}$ 可以采用冷酷触发策略来维持一个均衡，在该均衡中，每一轮$\mathcal{C}$ 都将剩余s 划分为（e^*,$s - e^*$）。

我们在附录B.3中给出该定理的证明，其中包括图7中策略组合的博弈树。我们的证明很直接，采用分析重复博弈的标准论证：我们评估合作路径上的收益以及$\mathcal{C}$ 偏离时的收益，并证明由于$\mathcal{A}$的冷酷触发威胁，$\mathcal{C}$ 在前者中获得的收益更高。

6 实证评估

在本节中，我们利用历史区块链数据，在真实场景下开展实验，展示抢先交易对植入后门的 CoinAlg 的影响。通过这些实验，我们了解了即使信息有限且不精确，抢先交易 CoinAlg 仍能带来多大利润。所有实验的源代码均可在此处找到：https://git.disroot.org/verge24/calgs。

动机。 我们在第 5 节展示了 CoinAlg 中利润攫取的理论风险。现在，我们旨在探究这种攫取在真实市场条件下可能达到的相对规模。CoinAlg 是否确实面临风险？抢先交易者能否利用有限的信息获利？

方法概述。 沿用第 5 节中的模型，我们设计了一种合成 CoinAlg，它能完美预知未来市场条件，并在历史区块数据上模拟其交易。它代表了理想的 CoinAlg：其交易以私有方式提交，只要它未被阻止交易，投资者就不会亏损。一个简单的抢先交易者试图利用它的交易获利。这些模拟反映了真实代币市场中的流动性条件和费用，同时也展示了即便是最理想情况下的 CoinAlg，也会面临抢先交易风险。

本节的实验补充了前文的理论结果，为以下观点提供证据：

> 关键洞见 3。 凭借准确的市场预测而获利的 CoinAlg，即使面对一个带宽很低、会向套利对手泄露交易信息的信道，也可能遭受显著的利润损失。

正如我们将在本节后文详细阐述的那样，我们使用一种专门的测试工具，开展了多种实验，以在反映以下因素的不同变量配置下检验这一假设：

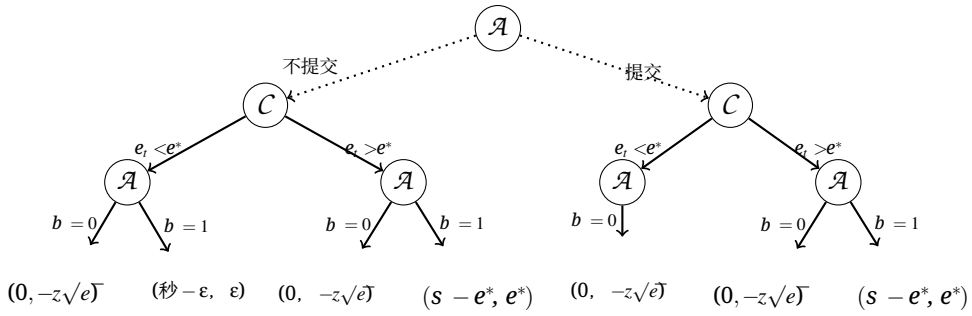

图 7：在博弈 $\mathcal{A}$ 中导致有利均衡的策略组合对应的 U-Sdwch 博弈树（图 6）。均衡位于最右侧节点。

现实世界中的考量：多种资产类型（稳定型和波动型）、CoinAlg 设计（透明和部分私有）、对抗策略（策略窃取和三明治攻击）等。总体而言，我们的实验表明，无论初始条件如何，只要 CoinAlg 泄露哪怕少量信息，就面临利润损失的风险。

6.1 实验设置与方法

CoinAlg 策略。 CoinAlg 按照预定时间表进行交易，并会使用尽可能多的资金，交易至其下一次预定交易时预测的未来市场价格，以确保获得利润。我们模拟的 CoinAlg 采用了比第 5 节博弈论模型中所体现的策略更为复杂的策略，后者为便于分析而进行了简化。该 CoinAlg 使用持续更新的信息：它在每个交易周期根据下一次预测价格调整其策略，并不一定会卖出上次交易中买入的资产。例如，如果 CoinAlg 预测以太币今天的价格将上涨 50 美元，它会尝试立即买入以太币；如果 CoinAlg 在次日预测价格还将上涨 70 美元，它会尝试在次日买入更多以太币，而不会卖出当前持有的以太币。它通过私有交易中继提交交易，以防范普通的三明治攻击（即抢先交易者在交易被纳入区块之前无法观察到其交易）。

我们实验的核心目标之一，是了解 CoinAlg 的利润损失在多大程度上取决于泄露的信息量：即使泄露的信息很少，是否也足以实现价值提取？因此，我们针对代表不同泄露类型的几种变体开展实验：(1) 采用固定间隔进行交易的完全透明交易策略；(2) 按固定时间表进行交易、仅向抢先交易者泄露二元信号（“买入/卖出”）的私有策略；以及 (3) 交易时间保密、但交易时间分布已知的透明策略。具体而言，对于第三种类型，CoinAlg 根据泊松过程安排交易，目标是平均每天交易两次（$\lambda = 3$，600 个区块）。这三种设置旨在检验几种可能的信息泄露程度会如何影响 CoinAlg 的盈利能力。

以这三种 CoinAlg 类型为起点，我们进一步调整初始条件中的底层细节，以检验这些因素是否也会影响我们的结果。具体而言，我们尝试了不同的资产类型、初始余额、交易执行时段，以及 CoinAlg 运行的总时间窗口。参数变化情况及其取值汇总见图8 。我们的目标是广泛考察各种 CoinAlg 类型：交易波动性资产与稳定资产的类型、余额较大与较小的类型、运行时间较短与较长的类型，等等。正如我们将看到的，调整这些参数并不会改变结果的方向。

对抗策略。谈到对手，我们同样考虑了多种策略，这些策略受到前文所述威胁的启发：(1) 策略窃取，即抢先交易者复制 CoinAlg 的交易策略；(2) 三明治攻击，即对手怀疑 CoinAlg 策略的盈利能力，于是对 CoinAlg 的交易进行抢先交易和尾随交易。

对于三明治攻击，一个重要的考量是对手应在其交易中使用的资金量。虽然人们可能会认为，最佳策略是尽可能使用所有可用资金，但这并非最优，原因有二：（1）规模更大的交易会产生更高的资金池费用；（2）规模更大的交易可能导致资产价格上涨至超过 CoinAlg 的价格上限。因此，给定某一笔需要抢先交易的 CoinAlg 交易，对手估算的三明治攻击利润实际上是其交易输入的单峰函数，其峰值由上述两个条件决定。我们在图9中展示了一个示例，量化了对手利润与其三明治攻击输入规模之间的关系。

由于该函数严格单峰，我们可以对抢先交易者的交易规模进行黄金分割搜索，以优化抢先交易者的交易输入 [18]。我们使用这一技术，根据抢先交易者可用的资金推导其三明治攻击交易。

CoinAlg 类型	对抗策略	资产	CoinAlg 初始余额	攻击者初始余额	开始时间偏移	执行窗口
• 完全透明 • 透明、秘密交易时间 • 带有一位信息的私有渠道	• 策略窃取 • 夹击	• USDC • PEPE	• 12以太币 • 60以太币 • 300以太币 • 1,500 以太币	• 8以太币 / 16 ,000 USDC • 40以太币 / 80,000 USDC • 200以太币 / 400 ,000 USDC • 1,000 以太币 / 2,000,000 USDC	• 0小时 • 6小时 • 12小时 • 18小时	• 0.75M 个区块 • 1.5M 个区块 • 2.25M 个区块 • 3M 个区块

图 8：实验参数及其取值汇总。测试框架会评估所有取值的笛卡尔积。

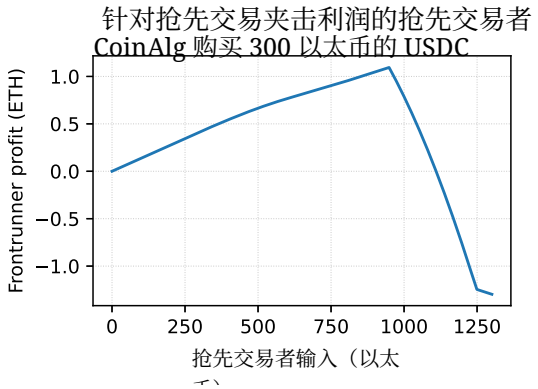

图9：在以太坊区块 20,000,000 的 0.05% Uniswap V3 池中，对 CoinAlg 以 300 以太币购买 USDC 的交易进行三明治攻击所获利润，随抢先交易者的交易规模而变化。当 CoinAlg 开始减少投入资金、遵守其价格上限时，利润便会下降。若抢先交易者投入 1,250 以太币，CoinAlg 不会发起交易。

与 CoinAlg 类似，我们针对对手的初始余额考虑多个取值，以分析资金充足和资金不足的对手提取利润的可行性。参见图8 ，其中总结了我们考虑的各个取值。正如我们稍后将讨论的那样，我们的模拟表明，这两类对手都能够获利。

模拟细节。我们在两个真实的 Uniswap V3 代币池（USDC/ETH，手续费为 0.05%；PEPE/ETH，手续费为 0.3%）上进行实验，并将交易周期设定为以太坊区块 20,000,000 至 23,000,000 之间（2024年6月1日至2025年7月26日）。我们使用归档节点模拟交易：每当参与者想要提交交易时，便在目标区块的开头运行一次多交易调用。该调用（1）通过从销毁地址转移资产（以太币、PEPE），或使用获授权的代币铸造者（USDC），为 CoinAlg 和抢先交易者的模拟账户注入资金；（2）在目标市场上授权并兑换代币；（3）返回交易结果，并报告参与者的最终余额

。

为了防止这些交易扰乱模拟中的历史交易活动，每个模拟区块结束后都会恢复历史状态。我们加入了第三个“幽灵”参与者，在每次多调用开始时进行一笔交易，以抵消此前区块中 CoinAlg 和抢先交易者造成的市场变动。在 CoinAlg 完成交易且抢先交易者有机会对此作出反应后，我们会消除这些市场修正。这样可以确保 CoinAlg 的价格预测保持准确，并使代币价格易于解读。请注意，这会抵消抢先交易者回跑交易的影响：回跑交易不会对 CoinAlg 造成负面影响。因此，我们的实验结果展示了 CoinAlg 利润的上限；抢先交易的影响可能更大。我们的模拟与模型还有一点不同：CoinAlg 获得的资金是有限的。

请注意，对于一组固定的初始条件（即从图中每列选取一个值8），模拟运行结果是确定性的。因此，每组参数只需运行一次即可得出结果。唯一的例外是 CoinAlg 按照泊松分布决定何时进行交易的实验。在这种情况下，对于每种参数组合，我们使用不同的随机种子进行了 16 次实验运行。

6.2 结果

我们使用测试框架，对图中所示的全部 3,072 种参数组合进行了测试8。所有实验运行的输出均可在此处找到：
https://git.disroot.org/verge24/calgs/src/branch/main/results。

我们的实验得出的主要结论是，在所有参数组合中，存在对手都会导致 *CoinAlg* 利润损失，这证实了我们的假设：泄露信息（无论是向公众泄露，还是通过后门向内部人员泄露）的 CoinAlg 都要付出代价。虽然利润损失的具体数额取决于初始条件（尤其是初始余额），但所有情况下的变化趋势都一致。

现在，我们来探讨由……揭示的一些更底层的启示

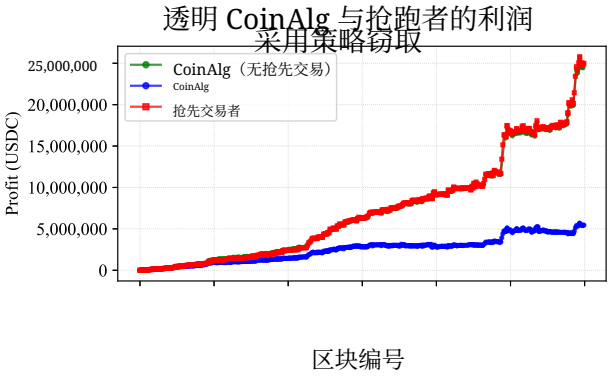

图10：实验运行结果，展示了理想 CoinAlg（蓝色）使用透明交易算法在 WETH/USDC 0.05% 池中以 300 WETH 起始时的利润、采用策略窃取的抢先交易者（红色）以 200 WETH / 400,000 USDC 起始时的利润，以及没有抢先交易者时同一 CoinAlg（绿色）的利润。请注意，图中抢先交易者的利润与没有抢先交易时 CoinAlg 的利润几乎相同。

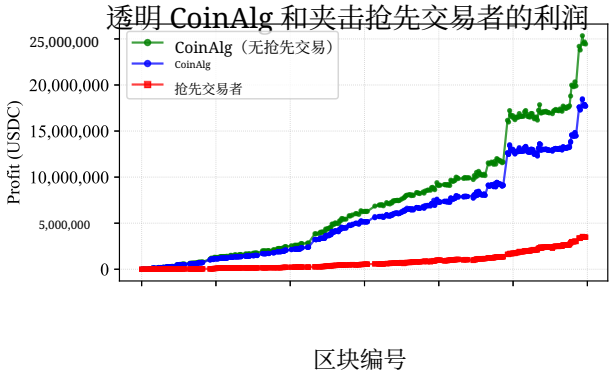

图11：实验运行结果，展示了透明 CoinAlg（蓝色）、夹击式抢跑者（红色），以及在没有抢跑者时的 CoinAlg（绿色）的利润；实验配置与图10相同。

我们的实验。

两种对抗策略都切实可行。我们的实验证实，正如预期的那样，策略窃取和三明治攻击对对手而言都是有利可图的策略。

对于策略窃取，抢先交易者会复制 CoinAlg 的每笔交易，并在 CoinAlg 提交交易前一个区块提交自己的交易，从而获得更好的价格，并减少价格上限范围内剩余的可用流动性。图10展示了我们一次实验运行的结果，该实验考察了透明 CoinAlg 和实施策略窃取的对手。如我们所见，抢先交易者几乎攫取了 CoinAlg 在没有抢先交易时本可获得的全部利润。资金充足时，其规模更大的市场交易会耗尽可用流动性并触及目标价格上限，导致抢先交易者的利润波动更大，同时显著削弱 CoinAlg 的交易能力。当抢先交易者能够将价格推至价格上限时，CoinAlg 就无法进行交易。

传统夹心攻击活动通常会在同一区块中围绕受害者交易安排头寸交易。采用防抢先交易最佳实践（例如私有内存池）的 CoinAlgs，如果其交易可以被预测，仍可能遭受三明治攻击。正如图 11所示，抢先交易者发起的三明治攻击使我们的透明CoinAlg利润减少了 27.6%。尽管理想的 CoinAlg 仍会受到三明治攻击活动的影响，但它依然能够盈利，因为其使用的价格上限可确保交易有利可图，且交易并非基于相对于市场价格的“滑点”；此外，当 CoinAlg 选择持有资产而非交易时，抢先交易者无法影响它。

即使是部分私有的CoinAlgs也容易受到攻击。我们的实验还表明，对手无需完全了解 CoinAlg 也能攫取利润，即使信息有限，也可能造成价值损失。

我们通过修改 CoinAlg 来研究这种情形，使其交易时间保密：它不再按照固定时间表进行交易，而是根据泊松过程执行交易。与传统的三明治攻击不同，抢先交易者现在必须承担风险，因为它会在较长时间内持有 CoinAlg 的头寸。我们为抢先交易者设计了一个简单的算法：一看到 CoinAlg 的交易，它就立即计算 CoinAlg 下一次交易的预期时间（未来 3,600 个区块），并尝试抢先执行一笔由 CoinAlg 算法针对该预期目标区块生成的交易。CoinAlg 最终完成交易后，抢先交易者便提交其跟随交易。最终，如图12所示，抢先交易者仍然获利颇丰。

我们进一步考虑能力更加有限的对手：此时 CoinAlg 为私有状态，但对手可以访问一个低比特率的隐蔽信道。我们不详细说明该信道的构造，但认为它可以通过与 CoinAlg 处理过程相关的加密定时侧信道等方式实现。在 CoinAlg 下一次交易时间前十个区块，该信道会向抢先交易者传送一个比特，告知其下一笔交易的预期方向。该信道不会传递确切的交易金额或价格上限，甚至不会告知 CoinAlg 是否会提交交易——如果交易无利可图（例如，市场价格与其限价过于接近），CoinAlg 可能不会提交交易。抢先交易者针对 CoinAlg 采取一种温和的三明治攻击策略，盲目地假设 CoinAlg 会造成固定的 1% 价格影响。

如图所示 13，在掌握这些有限信息的情况下，抢先交易者仍然能够获利，并使CoinAlg的利润减少17.8%。尽管抢先交易者的若干笔个别交

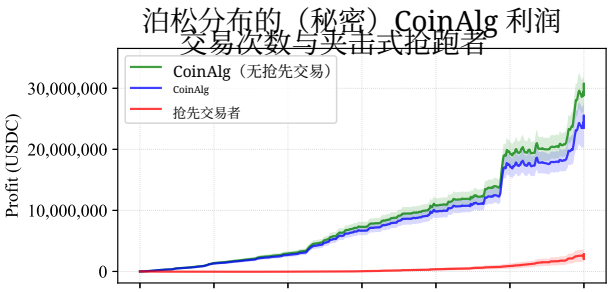

图 12：CoinAlg（蓝色）从 300 WETH 初始资金开始，使用随机化的隐藏交易时间进行 WETH/USDC 交易；抢先交易者（红色）从 200 WETH / 400,000 USDC 初始资金开始，实施远程三明治攻击；以及不存在抢先交易者时同一 CoinAlg（绿色）的 256 次运行平均利润。色带显示第 10 至第 90 百分位数的范围。

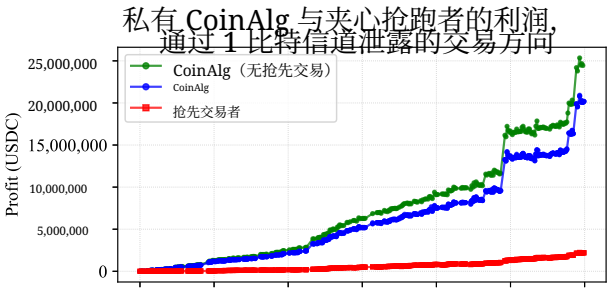

图 13：私有 CoinAlg（蓝色）进行 WETH/USDC 交易，仅向实施三明治攻击的特权抢跑者（红色）泄露交易方向，并估算每笔交易造成 1% 的价格变动；以及不存在抢先交易者时同一 CoinAlg（绿色）的利润对比。

易并未获利，但其成功交易带来的收益足以弥补亏损。这名抢先交易者对CoinAlg利润的影响，超过了抢先交易者对完全透明CoinAlg影响的一半。

市场影响较小的集体交易不受抢先交易影响。在我们的实验中，当交易给抢先交易者造成的成本超过其从三明治交易中可获取的价值时，抢先交易者就会选择不进行交易。例如，对于费率为 0.05% 的 USDC/WETH 池，在流动性充足的池中实施三明治攻击，成本至少为输入金额的 $0.05\% \times 2 = 0.1\%$。当可用流动性相对于 CoinAlg 的交易规模较高时，这项成本会占据主导地位，正如 [35]中也观察到的。

7 启发式防护措施

我们的研究结果表明，CoinAlg 困境无法避免；任何缓解措施都无法彻底解决这一问题。然而，CoinAlgs（尤其是私有CoinAlgs）似乎注定会成为投资格局的一部分。因此，找到能够限制私有 CoinAlgs 中 CoinAlg 困境影响的防护措施至关重要。本节将简要介绍几种实用的启发式防护措施，希望能为未来的研究提供启发。

随机化包装器。 对 *CoinAlg* $\mathcal{C}$ 的交易进行随机化，可以增加对手从中榨取价值的难度。具体而言，$\mathcal{C}$ 可以使用由一种透明 机制生成的私有随机性来选择下一笔交易 u，该机制确保公平性，即不会向对手泄露信息，也不会受对手影响。我们称这种机制为随机化包装器。

随机化包装器将私有 CoinAlg 与透明的防护措施结合起来，既保留私有交易策略带来的盈利能力优势，又限制内部人员通过信息泄露进行利用的机会。

回顾第3 节，$\mathcal{C}$ 通过以下步骤在状态$\mathcal{S}$中选择交易：(1) 运行交易生成算法 $\pi_{\mathcal{S}} \leftarrow \mathsf{ALG}_{\mathsf{dist}}(\mathcal{S})$；(2) 运行采样算法 $\tau \leftarrow\!\$\, \mathsf{ALG}^{\mathsf{sample}}(\pi_{\mathcal{S}})$。我们的想法是~替换步骤 (2)，并用以下两个组件实现新的私有 CoinAlg $\mathcal{C}$：

1. 一个私有 算法 $\pi_{\mathcal{S}} \leftarrow \mathsf{ALG}_{\mathsf{dist}}(\mathcal{S})$，以及
2. 一个透明的随机化包装器RW。

这里，$\mathsf{RW} = (\mathsf{RW}\$, \mathsf{RW}_e)$，其中 $\tau \leftarrow\!\$\, \mathsf{RW}\$(\pi_{\mathcal{S}})$ 使用公平的私有随机币选择一笔交易，而 RW_e 随后执行τ。RW 之所以透明，是因为其代码已公开发布，其属性可以公开验证。（不过，其状态在执行期间并不透明。）通过在可信执行环境中运行$\mathcal{C}$，可以确保$\mathsf{ALG}_{\mathsf{dist}}(\mathcal{S})$及交易分布 $\pi_{\mathcal{S}}$的隐私。同时，可信执行环境可以确保RW 正确执行，并使用由可信执行环境生成的公平、保密的随机币。

总而言之，一种私有的随机化交易策略，用户可以公开验证RW，以确保它使用公平的随机性选择交易，且不会将随机性泄露给对手。虽然随机化包装器无法彻底消除攻击，但有助于最大限度地减少价值提取。

我们注意到，在某些情况下，随机化包装器可能会导致 CoinAlg 执行次优交易，因为包装器可能会抽取一笔并非利润最高的交易。然而，也存在随机化不会牺牲利润的情况。例如，如果某个 CoinAlg 正在进行长期预测，包装器可以将交易发生的时间随机化，从而有效地充当时序抖动。在这种情况下，对手将无法利用有利可图的三明治攻击所需的瞬时价格上涨。

受保护的训练管道。 随机化包装器会修改 *CoinAlg* 的运行时 行为，以尽量减少价值提取。另一种方法是在 CoinAlg 的 创建 阶段对其进行修改。具体而言，如果用于训练基于人工智能的 CoinAlg 的 数据源是公开的，这将确保模型在训练过程中未被以恶意方式设计。不过，即使数据源列表是透明的，仍有一个关键技术挑战：如何保证模型确实（仅）使用了来自这些数据源的数据进行训练。为此，可以使用 受保护的管道 [16] 对 CoinAlg 进行训练，这些管道能够安全地证明人工智能模型所使用的训练数据源。

漏洞赏金系统。 除了技术性对策之外，还可以通过经济手段，即设立漏洞赏金系统，来遏制对 CoinAlg 困境的利用。其理念是让用户能够证明价值提取后门的存在并领取金钱奖励，从而激励人们仔细审查 CoinAlgs，甚至进行举报。在 CoinAlg 漏洞的背景下，这类漏洞赏金系统可以做到公开可验证——甚至可以通过智能合约实现 [4]。申领者提交一系列关于 CoinAlg $\mathcal{C}$交易的正确预测。若该序列具有足够的预测熵，便构成了能够访问隐蔽（或侧）信道的可靠证明。

作为一个简单、具体的例子，考虑上述构造中用于防止随机化包装器RW 发生交易信息泄露的赏金系统$\mathcal{B}$。假设证明者／举报者P 声称，在$\mathcal{C}$执行任何交易 τ 之前，就能获知一个谓词$F(\tau)$。证明系统的工作方式如下。P向$\mathcal{B}$发送F。随后，在一系列纪元$i \in [1,t]$ 中的每一个纪元里，P 都会在执行 τ_i 之前，向$\mathcal{B}$ 发送对$F(\tau_i)$ 的预测x_i。在纪元$t+1$，$\mathcal{B}$ 会检查两个条件：

1. *Predictions are correct:* $\forall i \in [1,t] : x_i = F(\tau_i)$.
2. 已达到可靠性阈值：预测序列 $\{x_i\}$可为系统参数 θ（例如 $\theta = 128$）生成可靠性为 $1-2^{\theta}$ 的证明。

令 $X_i = F(\tau_i \leftarrow\!\$ \mathsf{RW}(\pi_{S_i}))$ 表示关于RW$的私有随机数的随机变量。为验证条件 (2)，$\mathcal{B}$ 将 (F,t,θ) 传递给集成到RW 中的应用程序接口，以支持赏金系统；该接口会检查：

$$\sum_{i=1}^{t} H_\infty(X_i) \geq \theta.$$

8 相关工作

CoinAlgs。 尽管没有使用这一称谓，但传统金融中的 CoinAlgs 已得到广泛研究，因为它们涵盖了各种金融工具（见第 2 节）。例如，既有研究考察了有效市场假说（即仅凭公开信息不可能实现跑赢市场的表现 [23, 29]）；不透明金融工具带来的风险则促使人们研究投资者保护和受托责任 [19, 28], ，包括针对机器人顾问的相关研究 [14]。然而，尚无既有研究同时探讨隐私与透明性，或二者之间固有的张力。

关于 Web3 CoinAlg 的文献有限，且主要聚焦于人工智能驱动的去中心化自治组织。值得注意的是，Patlan 等人 [25] 对 ElizaOS 进行了安全性分析，并揭示了提示注入漏洞。这些研究结果与我们的工作关注点不同且相辅相成，因为它们展现了人工智能驱动的去中心化自治组织面临的其他问题。

三明治攻击。此前已有大量研究探讨三明治攻击。与我们的工作最相关的是，有几篇论文像我们在第5节中一样，从博弈论角度分析三明治攻击。Heimbach 和 Wattenhofer [13] 提出了三明治博弈，以对这类攻击进行形式化分析，但关键区别在于，他们的分析是一次性的，而我们考虑的是无限重复变体。Zhou 等人 [35] 也对三明治攻击进行了形式化描述，但其分析仅限于攻击者，忽略了受害者一方的考量，例如滑点限额调整。其他研究通过实证方法研究三明治攻击，与第6节类似。许多论文 [27, 30, 35] 量化了通过三明治攻击及其他最大可提取价值攻击获取的价值；而我们的实验引入了新的市场参与者，而非仅观察历史数据，因为私有 CoinAlg 无法被观察到，真正透明的 CoinAlg 也尚不存在。

重复博弈。 第 5 节结果的核心组成部分之一，是对我们三明治攻击重复变体的分析。去中心化金融文献中关于重复博弈的研究出人意料地少，因为大多数研究考察的是一次性或有限期博弈（例如 [13, 20]）。然而，去中心化金融涉及参与者之间反复进行的策略性互动，因此重复博弈具有重要的现实意义。少数研究确实考察了重复博弈。例如，Daian 等人 [8] 通过连续的自适应竞价对“优先级 Gas 拍卖”进行建模；虽然他们的分析针对的是一次性博弈，但每次迭代都类似于重复博弈。同样，Judmayer 等人 [15] 使用重复博弈对区块链资产进行建模，以探讨最大可提取价值定义。

加密货币与人工智能。 我们的论文为加密货币与人工智能这一新兴交叉领域作出了贡献，例如，关于用于人工智能计算的“去中心化物理基础设施网络” [21], 、面向智能体人工智能的支付 [24], 、“无需信任”的智能体间协议 [9],、经过认证的机器学习流水线 [16],等。加密货币与人工智能正迅速融合，而 CoinAlg 困境正是这一融合的核心，因为任何涉及多名用户的、基于人工智能的金融工具都可能适用于我们的研究结果。

9 结论

本研究开启了对 CoinAlg 困境的探讨：在一大类金融工具 CoinAlgs 中，盈利能力与经济公平性之间存在权衡。我们认为，CoinAlgs 要么保持透明，但面临利润受损的风险；要么保持私密，从而为内部人员进行不公平价值提取打开方便之门。

通过广泛的实验，我们表明，这种威胁并非只是理论上的担忧——即使在 CoinAlg 看似无害的情况下，也可能被植入隐蔽的后门，从而从对手身上获利。然而，无论这种根本性的矛盾如何，CoinAlgs 都是金融格局中不可避免的一部分，因此，寻找有原则的防护措施仍是未来研究的重要方向。

致谢

本研究由 NSF CNS-2112751 资助，并获得 IC3 产业合作伙伴和赞助商的大力支持。Ari Juels 担任 Chainlink Labs 首席科学家。

参考文献

[1] Kushal Babel、Philip Daian、Mahimna Kelkar 和 Ari Juels。《Clockwork finance：智能合约中经济安全性的自动化分析》。载于 *IEEE S&P*，2023 年。

[2] Kyle Bagwell。《Commitment and Observability in Games》。*Games andEconomicBehavior*，1995 年。

[3]Barbara Friedberg。《按资产管理规模排名的十大机器人顾问》。https://www.forbes.com/advisor/investing/top-robo-advisors-by-aum/，2024 年。

[4] Lorenz Breidenbach、Phil Daian、Florian Tramèr 和 Ari Juels。《Enter the hydra：迈向有原则的漏洞赏金与抗利用智能合约》。载于第 *27* 届 *USENIX Security Symposium*（*USENIX Security 18*），第 1335–1352 页，2018 年。

[5] Coinbase。AI16Z 价格。https://www.coinbase.com/price/ai16z，2025年。访问日期：2025-09-08。

[6] CoW Protocol Association。Cowswap。https://swap.cow.fi/，2025年。访问日期：2025-10-11。

[7] Crypto.com University。什么是 ElizaOS? https://crypto.com/es/university/what-is-elizaos，2025年。访问日期：2025-09-08。

[8] Philip Daian、Steven Goldfeder、Tyler Kell、Yunqi Li、Xueyuan Zhao、Iddo Bentov、Lorenz Breidenbach 和 Ari Juels。《Flash boys 2.0：去中心化交易所中的抢先交易、矿工可提取价值与共识不稳定性》。载于*IEEES&P*，2020 年。

[9] Marco De Rossi、Davide Crapis、Jordan Ellis 和 Erik Reppel。ERC-8004：无信任代理。https://eips.ethereum.org/EIPS/eip-8004，2025年。

[10]Flashbots。Flashbots Protect RPC。https://protectrpc.flashbots.net/，2025年。访问时间：2025- 10-11。

[11] Mathias Hall-Andersen 和 Nikolaj I Schwartzbach。区块链上的博弈论：智能合约游戏模型。发表于国际算法博弈论研讨会，2021年。

[12] John C Harsanyi。论合作博弈理论所依据的理性公理。《冲突解决杂志》，1961年。

[13] Lioba Heimbach 和 Roger Wattenhofer。借助博弈论消除夹心攻击。载于 *Asi-aCCS*，2022年。

[14] Megan Ji。机器人是合格的受托人吗？根据1940年《投资顾问法》监管机器人顾问。《哥伦比亚法律评论》，2017年。

[15] Aljosha Judmayer、Nicholas Stifter、Philipp Schindler 和 Edgar Weippl。估算（矿工）可提取价值很困难，那就去购物吧！载于国际金融密码学与数据安全会议，第74–92页。Springer，2022年。

[16] Ari Juels 和 Farinaz Koushanfar。《机器学习安全的道具》。*arXiv* 预印本 *arXiv:2410.20522*，2024 年。

[17] Adam Tauman Kalai、Ehud Kalai、Ehud Lehrer 和 Dov Samet。《承诺民间定理》。《博弈与经济行为》，2010 年。

[18] Jack Kiefer。《求最大值的序贯极小极大搜索》。《美国数学会会刊》，1953 年。

[19] Arthur B. Laby。《经纪交易商和投资顾问的信义义务》。《维拉诺瓦法律评论》，2010 年。

[20] Aron Laszka、Benjamin Johnson 和 Jens Grossklags。《比特币矿池何时会枯竭》。载于《金融密码学》，2015 年。

[21] Zhibin Lin、Taotao Wang、Long Shi、Shengli Zhang 和 Bin Cao。《去中心化物理基础设施网络（DePIN）：挑战与机遇》。《*IEEE*网络》，2024 年。

[22] George J Mailath 和 Larry Samuelson。《重复博弈与声誉：长期关系》。牛津大学出版社，2006年。

[23] Burton G Malkiel。《对有效市场假说三十年后的思考》。*FinancialReview*，2005年。

[24] Bill Marino 和 Ari Juels。赋予人工智能代理访问加密货币和智能合约的权限会带来新的人工智能危害途径。*arXiv* 预印本 *arXiv:2507.08249*，2025年。

[25] Atharv SinghPatlan、PeiyaoSheng、S Ashwin Hebbar、Prateek Mittal 和 Pramod Viswanath。Cryptoland 中的人工智能代理：实际攻击与没有一招制胜的良方。密码学电子印本档案，2025年。

[26]Pragmatic Coders。面向交易员的顶级人工智能工具。https://www.pragmaticcoders.com/blog/top-ai-tools-for-traders，2025年6月13日。访问日期：2025-09-08。

[27] Kaihua Qin、Liyi Zhou 和 Arthur Gervais。《量化区块链可提取价值：森林有多黑暗？》发表于*IEEES&P*，2022年。

[28] Yuki Sato。《金融市场中的不透明性》。《金融研究评论》，2014年。

[29] Allan Timmermann 和 Clive WJ Granger。《有效市场假说与预测》。《International Jour- nal of Forecasting》，2004年。

[30] Christof Ferreira Torres、Ramiro Camino 等。《抢先交易者 Jones 与黑暗森林掠夺者：以太坊区块链上抢先交易的实证研究》。发表于第*30*届*USENIX*安全研讨会 *(USENIX Security 21)*，第1343–1359页，2021年。

[31] Truth Terminal。在 X 上的 Truth Terminal。https://x.com/truth_terminal，2025年。访问日期：2025-09-08。

[32] Heinrich von Stackelberg。《市场结构与均衡》。Damien Bazin、Lynn Urch 和 Rowland Hill 译（Springer，2011年）。

[33] John Werner。人工智能机器人凭借前所未有的股票收益惊艳众人。《福布斯》，2025年1月3日。访问日期：2025-09-08。

[34] Jeff Wilser。《Truth Terminal：人工智能加密货币的奇异未来》。https://www.coindesk.com/tech/2024/12/10/the-truth-terminal-ai-crypto-s-weird-future，2024年。发布日期：2024年12月10日。访问日期：2025-09-08。

[35] Liyi Zhou、Kaihua Qin、Christof Ferreira Torres、Duc V Le 和 Arthur Gervais。去中心化链上交易所中的高频交易。载于*IEEE S&P*，2021年。

[36] Liyi Zhou、Xihan Xiong、Jens Ernstberger、Stefanos Chaliasos、Zhipeng Wang、Ye Wang、Kaihua Qin、Roger Wattenhofer、Dawn Song 和 Arthur Gervais。《综述：去中心化金融（DeFi）攻击》。发表于 *IEEE S&P*，2023 年。

定理 1 的证明1

证明。回顾定义1 指出，对于任意状态集合$\mathbb{S}$，分布 π_S（$\mathcal{C}$从中抽取交易）和 $\pi_{\mathcal{LS}}$（由公开函数诱导的、来自$\mathcal{C}$的可能交易）的平均统计距离至少为 ε。对此定义的一种等价表述是存在一个区分器 Dist，它能够以至少 ε 的概率区分来自这两个分布的样本。即：

$$\left|\Pr\left[x \leftarrow\$ \pi_{\mathcal{S}} : \mathsf{Dist}(x) = 1\right] - \Pr\left[x \leftarrow\$ \pi_{\mathcal{S}}^{\mathcal{L}} : \mathsf{Dist}(x) = 1\right]\right| \geq \varepsilon$$

因此，我们的目标是证明这个区分器存在。直观地说，Dist 将利用以下事实：根据经济不公平性的定义，存在一种对抗算法，在使用完美预测预言机 O（即接收来自 π_S 的样本）时提取的价值，比任何改为接收来自 π_S 的样本的其他算法都多。

根据假设，由于存在 $\alpha,t > 0$，使得相对于 O，C 是 (α,t)-不公平的，因此存在一个算法 A，使得对于每个可能的高效算法 P，A^O 提取的额外价值单位α 比 PL 多，且该情况发生的概率为 t。固定某个 P，例如使用 L 获得最大价值的算法。Dist 将以 A 和 P 为输入运行 FAIR 博弈，并使用其输入分布 $\tilde{\pi}$（它试图对该分布进行区分）来响应两个参与者的查询。具体而言，每当 A 发出查询时，Dist 都会抽取一笔交易 $x \leftarrow\$ \tilde{\pi}$ 并将 x 返回给 A。注意，当且仅当 $\tilde{\pi} = \pi S$ 时，这与 A 从 O 获得的分布相同。接着，每当 P 发出查询时，Dist 会再次抽取 $x \leftarrow\$ \tilde{\pi}$，但这次会将公开函数 L 向 P 揭示的 x 的属性返回给 P。注意，无论 $\tilde{\pi}$ 的值如何，这与 P 从其预言机获得的分布相同。最后，如果在博弈结束时 A 提取的价值超过 α，Dist 就输出 1。根据假设，当 Dist 获得来自 πS 的样本时，其输出 1 的概率为 t。反之，当其获得来自 πLS 的样本时，输出 1 的概率为 0。由此便得到了我们所需的结果。

□

B 透明CoinAlgs结果的证明

B.1 引理1的证明

证明。初始自动做市商的储备金为$\mathsf{bal}_0^{\mathsf{AMM}} = [pr_0$ 美元，r_0 TOK$]$。将 TOK 价格提升至$p + p_{\mathsf{lim}}$ 后，新的余额为 $\mathsf{bal}_1^{\mathsf{AMM}} = [(p + p_{\mathsf{lim}})r_1$ 美元，r_1 TOK$]$。由于这是一个恒定乘积自动做市商，我们有：

$$(p + p_{\mathsf{lim}})r_1^2 = pr_0^2.$$

因此，

$$r_1 = r_0\sqrt{\frac{p}{p + p_{\mathsf{lim}}}}.$$

USD 交易成本为：

$$\begin{aligned}\mathsf{bal}_1^{\mathsf{AMM}}(\mathrm{USD}) - \mathsf{bal}_0^{\mathsf{AMM}}(\mathrm{USD}) &= (p + p_{\mathsf{lim}})\, r_1 - p\, r_0 \\ &= (p + p_{\mathsf{lim}})\, r_0 \sqrt{\frac{p}{p + p_{\mathsf{lim}}}} - p\, r_0 \\ &= p\, r_0 \left(\sqrt{\frac{p + p_{\mathsf{lim}}}{p}} - 1 \right). \end{aligned} \tag{2}$$

□

B.2 引理2的证明

证明。

$$\begin{aligned} p\, r_0 \left(\sqrt{\frac{2p+e}{p}} - 1 \right) &\le p\, r_0 \sqrt{\frac{2p+e}{p}} \\ &\le p\, r_0 \left(\sqrt{\frac{2p}{p}} + \sqrt{\frac{e}{p}} \right) \\ &= p\, r_0 \sqrt{2} + \frac{p}{\sqrt{p}} \sqrt{e} \\ &= p\, r_0 \sqrt{2} + \sqrt{p}\sqrt{e} \\ &\le p\, r_0 \sqrt{2} + p\, r_0 \sqrt{2} \sqrt{e} \\ &\le p\, r_0 \sqrt{2} \sqrt{e} + p\, r_0 \sqrt{2} \sqrt{e} \\ &= z\sqrt{e}, \text{ where } z = p\, r_0 2\sqrt{2} \end{aligned}$$

□

B.3 定理4的证明

证明。 令 $0 < e^* \le s$ 为某个固定的截止点。考虑以下策略组合 Σ:

1. **去中心化自治组织 D:** 在每个回合 t，将 s 划分为 $(e_t, s - e_t)$，其中 $e_t = e^*$，前提是此前所有回合均满足 $b = 1$；否则，令 $e_t = \varepsilon$ 取某个最小值 ε。2. **套利者 $\mathcal{A}$:** 如果回合 t 满足 $e_t = e^*$，则输出 $b = 1$。可信地承诺 在 $e_t < e^*$ 之后的所有回合中输出 $b = 0$。

让我们分析这一策略组合的收益。沿着合作路径，$\mathcal{C}$ 和 $\mathcal{A}$ 的收益流为:

$$\Pi_{\mathcal{C}} = \sum_{t=0}^{\infty} c_1 (s - e_t) \delta_{\mathcal{C}}^t = \sum_{t=0}^{\infty} c_1 (s - e^*) \delta_{\mathcal{C}}^i = \frac{c_1 (s - e^*)}{1 - \delta_{\mathcal{C}}}$$

$$\Pi_{\mathcal{A}} = \sum_{t=0}^{\infty} c_2 e_t \delta_{\mathcal{A}}^i = \sum_{t=0}^{\infty} c_2 e^* \delta_{\mathcal{A}}^i = \frac{c_2 e^*}{1 - \delta_{\mathcal{A}}}$$

相反，假设 $\mathcal{C}$ 在某一回合 $\tilde{t}$ 偏离合作路径，即选择 $e_{\tilde{t}} < e^*$。若如此，$\mathcal{C}$ 在此后所有回合中的收益均为 0，因此其总收益严格低于沿合作路径获得的收益。重要的是，通过承诺在发生背叛时采用冷酷触发策略，$\mathcal{A}$ 将惩罚 $\mathcal{C}$，尽管这样做需要付出代价：如果没有这项承诺，冷酷触发策略就不是可信威胁，因此接受 $\mathcal{C}$ 的背叛实际上才是理性的。因此，$\mathcal{C}$ 会如愿留在合作路径上。

□