arXiv:2302.06708v2 [cs.CR] 7 Mar 2023

去中心化金融和非同质化代币阻碍区块链可扩展性

Lioba Heimbach、Quentin Kniep、Yann Vonlanthen、Roger Wattenhofer

苏黎世联邦理工
学院，瑞士苏黎世

摘要。 众所周知，许多经典区块链的交易吞吐量低得惊人，甚至低至比特币臭名昭著的每秒七笔交易上限。在区块链研究的第一个十年里，出现了各种提高吞吐量的方案和实现。但并发性究竟能达到多高？在早期，区块链主要用于用户之间的简单转账。然而，近来，去中心化金融（*DeFi*）和*NFT*市场彻底改变了区块链上的活动。二者均基于智能合约构建，并且广受欢迎。去中心化金融和NFT市场上的交易常常会与相同的智能合约交互。我们认为，这一发展改变了区块链的使用方式。在我们的研究中，我们对以太坊的交易图进行了历史分析，考察交易之间过去和现在的交互程度。我们发现，去中心化金融和NFT市场的兴起导致交易图中的“中心化”程度上升。如今，交易之间的相互关联更多：目前每个区块约有200笔交易，彼此之间存在4000条相互依赖关系。我们还发现，以太坊当前相互关联的交易工作负载可并行化程度有限。目前，实现超过五倍的加速比并不现实。

关键词： 区块链 · 以太坊 · 智能合约 · 去中心化金融 · 可并行化性 · 连通性 · 交易图。

1 引言

首个区块链比特币于 2008 年推出 [33]„ 使人们无需依赖中央机构即可执行金融交易。比特币的这一愿景促成了更多加密货币的诞生，其中最著名的是以太坊 [43], ，它于 2015 年引入了智能合约。然而，尽管加密货币持续走红，最受欢迎的[1] 加密货币的交易吞吐量仍然极低。

鉴于区块链的吞吐量较低，尤其是与 Visa 或 PayPal 等成熟支付系统相比，许多旨在解决区块链吞吐量低下问题的建议已被提出并付诸实施。

[1] 以用户愿意为使用区块链支付的总费用衡量（参见 https://cryptofees.info），以太坊的受欢迎程度远超其他支持智能合约的区块链，例如 Avalanche 和 Cardano。

二层协议 [13,36,25,34], 在链下处理交易、分片协议 [19,23,30,46,12,42,3],以及从工作量证明（*PoW*）转向权益证明（*PoS*） [18] 都是最广泛采用的扩容方案。除了开发上述解决方案外，多线程执行中并发控制的潜力也得到了深入探索。

然而，这些解决方案并未关注区块链交易性质变化所带来的影响。在去中心化金融（*DeFi*）和 NFT 市场迅速兴起之前，交易大多是双方之间的简单交易。因此，一个区块中大量交易之间很少存在依赖关系。由于依赖关系较少，所提出的解决方案可以通过并行执行交易来提高吞吐量。然而，去中心化金融和非同质化代币市场在提升吞吐量方面带来了新的挑战。

去中心化金融利用部署在区块链上的智能合约提供传统金融的许多服务，而 NFT 市场则利用智能合约促进 NFT 购买。构成去中心化金融和 NFT 市场核心的智能合约涉及区块中的许多交易，并在相当大比例的交易之间形成依赖关系。以太坊上的这一新现实给交易执行的并行化带来了巨大挑战。

在这项工作中，我们探究以太坊交易并行化的局限性。我们通过考察以太坊主网交易图随时间变化的连通性来分析这些局限。识别区块交易图中按所需执行工作负载衡量的最大连通分量和团，有助于我们探索并发执行潜力随时间的变化。具体而言，我们指出，去中心化金融中最重要的智能合约在交易图中处于核心位置，并造成了绝大多数交易依赖关系。因此，少数智能合约就构成了显著的并行化瓶颈；尤其是自 2020年 起，去中心化金融和非同质化代币市场得到了广泛采用。

这一发展给实现成熟支付系统的吞吐量水平带来了巨大挑战。因此，我们最后提出了三个需要着手解决的领域，以提高以太坊工作负载的可并行化性，并使并发机制充分发挥潜力。这些领域不仅应是我们分析重点——以太坊——需要关注的，也应是采用类似智能合约设计和使用模式的区块链需要关注的，因为我们预计它们会面临相似的瓶颈。

2 相关工作

区块链吞吐量一直是比特币研究最早关注的主题之一，针对这一问题已涌现出许多解决方案，例如第二层和分片。本文聚焦于通过并发执行实现工作负载并行化的解决方案。这些研究直接探讨工作负载的并行化，而我们则对其上限进行量化。

Sergey 和 Hobor [39] 是最早探索智能合约并发性、以实现区块链交易并行执行的研究者之一。他们将其与并发编程进行类比

。在 Zhang and Zhang [48], 提出的方案中，矿工使用并发控制技术预先计算可串行化调度，验证者重放区块时可利用该调度。Amiri 等人 [1]采用基于依赖图的并发控制技术，找到一种有效的执行调度，使互不冲突的交易能够并行执行。另一方面，我们的工作探索区块链交易依赖图，以量化现实世界中现有的并行化潜力。

此外，近期一系列围绕智能合约并发性的研究采用了推测执行。Dickerson 等人 [15],和 Anjana 等人 [2]通过推测执行，预先为区块中的交易计算出可串行化的并发调度；而 Gelashvili 等人 [21] 提出了 Block-STM，这是一种无需预先计算的并行执行引擎。Saraph 和 Herlihy [38]估算了矿工通过推测执行所能实现的潜在并发性。Chen 等人 [11] 则将推测执行提升到一个新层次：对等待纳入区块的交易进行推测执行。相比之下，鉴于去中心化金融和 NFT 市场近期兴起，我们探讨区块链交易的特性所决定的并发性极限。

Pîrlea 等人 [35] 和 Murgia 等人 [32] 利用静态分析实现并行执行。他们通过静态分析确定哪些交易可以安全地并行执行，以及哪些合约可以部署在不同的分片上。静态分析有助于提前识别依赖关系，但现有方法无法消除工作负载中固有的依赖关系，而这正是我们研究发现的核心。

一条并行的研究路线关注热门加密货币的交易图。Ron and Shamir [37] 最先分析了比特币交易图。虽然他们的工作以静态方式研究了完整的交易图，但 Kondor 等人 [26]也考察了比特币交易网络随时间发生的变化。另有多项研究通过时序图分析探索了以太坊交易网络 [22,31,31,10,4,20,29,28,49,24,45,44,49]。我们并非只研究以太坊完整交易网络的演变，而是关注智能合约使用量增加对以太坊逐区块交易图连通性的影响，并量化其对当前及历史交易工作负载可并行化性的影响。

在对以太坊交易网络的研究中，Zanelatto 等人 [47] 着重了解网络中连通分量的演变。他们的研究早于以太坊上去中心化金融和 NFT 市场的兴起，因此未能反映交易与不同智能合约之间日益增多的交互趋势。我们的工作聚焦于这种日益增强的连通性，并进一步讨论其对区块链的影响。

3 背景

下面将介绍有关以太坊区块链上的交易执行和 DeFi 智能合约的必要基础知识。

3.1 以太坊交易执行

以太坊是一个智能合约平台，也就是说，它不仅支持以太币转账。相反，它运行一个通用虚拟机，即 *以太坊虚拟机（EVM）*，该虚拟机执行特定的字节码指令集。因此，以太坊允许在交易中调用智能合约中定义的函数。智能合约函数可以调用其他智能合约的函数，从而生成内部调用。EVM 区分以下不同类型的函数调用：(1) 调用在被调用合约的存储范围内执行操作；(2) 委托调用在调用方合约的存储范围内执行操作；(3) 静态调用类似于常规调用，但仅对存储具有只读访问权限；(4) 代码调用是已弃用的委托调用版本。因此，所有调用都限定于特定合约，并且对其存储具有读写访问权限或只读访问权限。

如果两笔交易包含作用域相同的调用，且其中一笔交易对该作用域拥有写入权限，我们就称它们为冲突或依赖。请注意，与同一智能合约交互的两笔交易未必会访问同一个存储单元。不过，采用这种较为粗略的冲突判定方式，可以简化大规模分析，也更便于智能合约开发者理解。

此外，还有一些 EVM 指令（BALANCE、EXTCODESIZE、EXTCODE-HASH、EXTCODECOPY），允许合约读取任意地址的全局状态（包括当前以太币余额和智能合约代码），不受当前作用域限制。实际上，在这些指令中，主要是 EXTCODESIZE 被用于检测某个地址是否为智能合约。在我们的分析中，我们忽略了这些指令可能导致的冲突。

以太坊引入了访问列表 [7,8]，其中列出了交易希望访问的地址和存储键，并为这些访问提供燃气折扣。不过，访问列表是可选的，目前尚未得到广泛采用。我们发现，在超过 6 亿笔交易中，只有 200 万笔包含访问列表，而且其准确性和完整性尚不明确。构建仅包含地址的完整、准确的访问列表，对智能合约开发者和终端用户来说也会更容易。此外，与地址层级相比，先前的研究未能通过在存储键层级 [38]进行并行处理获得显著优势。

在 EVM 上执行代码需要支付*Gas*，费用会自动从发送方的以太币余额中扣除。Gas 用于衡量验证者在以太坊上执行代码的成本 [17]。因此，我们使用 Gas 作为实际执行时间（包括计算和存储访问）的代理指标。这样一来，便能尽可能减少对底层实现、运行时环境和硬件的假设。此外，我们将顺序 *Gas* 定义为并行调度中任何顺序执行所产生的最高累计 Gas 成本。

3.2 去中心化金融智能合约

去中心化金融提供了传统金融中的许多金融服务。去中心化金融不依赖中介机构，而是利用智能合约。下面将详细介绍去中心化金融中一些最重要的智能合约的功能。

ERC-20 代币是实现同质化代币的智能合约，即能够代表任何可拥有并以整数数量进行交换的事物，并遵循特定的接口标准 ERC-20（以太坊征求意见稿 20） [16]。这些合约至少实现一组规定的九个函数，包括 transfer （将一定数量的代币从调用者的地址转账到指定地址）和 balanceOf（获取指定地址的代币余额）。

路由器由去中心化交易所实现，作为其交易界面的统一前端。这些路由器是无状态的，只负责指定交易应通过哪些流动性池（ERC-20 代币的交易场所）进行路由。路由器会根据预定义的路线调用流动性池。

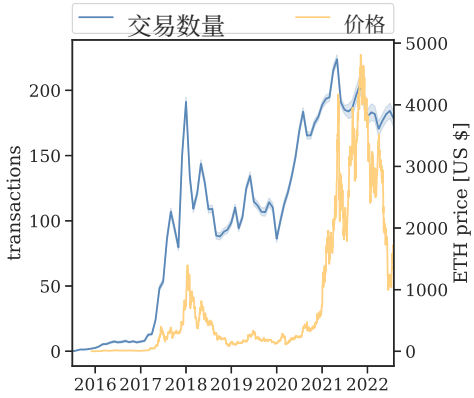
4 数据收集

我们运行 Erigon客户端 [27] 来收集以太坊区块链数据。具体而言，我们查询整个区块链历史的跟踪数据，以便更深入地了解以太坊主网工作负载的可并行化性及其随时间变化的趋势。跟踪数据为我们提供了每笔交易执行的内部调用。在接下来的评估中，我们通过随机抽取以太坊主网区块链整个历史期间每天的 65 个区块，来查看历史数据。通过历史分析，我们可以观察以太坊工作负载及其可并行化性的长期趋势。此外，我们还对近期数据进行了更细致的考察，分析了从 2022年6月1日到2022年8月31日这三个月期间的每一个区块。借助近期数据，我们希望准确了解以太坊区块链当前的可并行化状态。

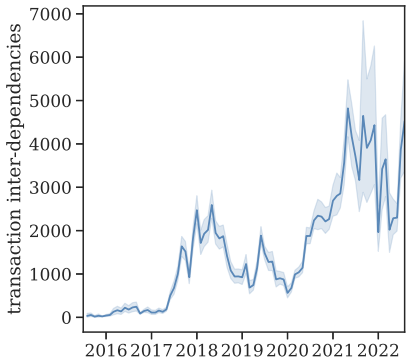
5 以太坊主网工作负载

接下来，我们将考察以太坊主网工作负载随时间的变化。在图 1a 中，我们按月计算并绘制数据集中区块的平均交易数量及其 95%置信区间。此外，我们还展示了以太币价格。

虽然最初每个区块中的交易数量很少（不到十笔），但从 2017 年初开始，每个区块中的交易数量首次显著上升，并于 2017 年末达到峰值，约为每个区块 200 笔交易。在此期间，加密货币市场出现了前所未有的投机热潮；这股热潮很可能推动了以太币价格和以太坊主网上交易数量的快速增长。随后市场热度消退，反映为每个区块的交易数量降至约 100 笔，以太币价格也大幅下跌。事实上，截至 2019 年底，每个区块的交易数量与以太币价格高度相关，皮尔逊相关系数为 0.77。之后，2020 年去中心化金融兴起，并引发了非同质化代币热潮，平均交易数量再度增加。此后，每个区块的平均交易数量稳定在约 200 笔，并未显著超过此前的峰值。随着去中心化金融和非同质化代币的发展，交易数量与以太币价格之间的相关性也有所下降；从 2020 年起，皮尔逊相关系数降至 0.60。



(a) 以太坊主网上每个区块的交易数量及以太币价格的历史变化。



(b) 每个区块的交易相互依赖关系数量的历史变化。

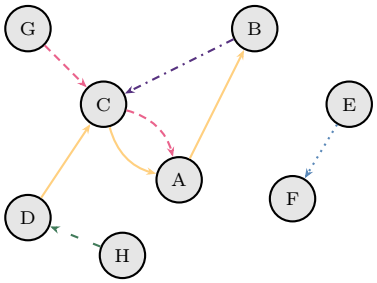
图1：每个区块的平均交易数量（参见图1a）和每个区块的平均交易相互依赖关系数量（参见图1b）的可视化。我们每天随机抽取65个区块，并绘制日平均值及95%置信区间。

我们还在图 1b 中绘制了交易相互依赖关系的数量，即访问同一智能合约的交易对数量。去中心化金融和 NFT市场对交易数量的影响虽未显著超过此前水平，但交易相互依赖关系的数量却并非如此。在最初的峰值期间，每个区块平均约有 2000 组交易相互依赖关系。值得注意的是，自 2020 年起，智能合约的使用量增加，使每个区块的平均交易相互依赖关系数量在峰值时升至 5000 组，此后始终未曾显著下降。请注意，这些相互依赖关系主要由少数核心 DeFi 智能合约产生，例如热门 ERC-20代币、去中心化交易所路由器和 NFT市场。因此，这些智能合约的广泛普及给以太坊交易并行化带来了重大挑战。

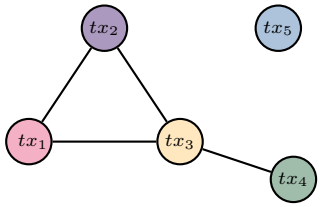
6 交易图表示

为了探究并行化潜力的趋势，我们考虑对交易数据进行图表示。以太坊交易数据通常采用两种图表示：基于地址的 和基于交易的。我们在定义 1 中给出了基于地址的图的定义，并在图 2a 中展示了一个示例。请注意，交易 tx_3 涉及四个地址： $D \to C \to A \to B$，而 tx_2 仅涉及两个地址： $B \to C$。

定义 1（基于地址的图（AG$_{n,\ m}$））。区块 $n, \ldots, m$, $n \leq m$ 的基于地址的图表示为 $AG_{n,\ m}(V, E, \{\omega_e\}_{e \in E})$。这里， V是图的顶点集合，每个 $v \in V$ 都是一个地址，而 V 是所有的集合



(a) 一组以太坊交易的基于地址的图表示，其中顶点为地址（合约或钱包），边表示调用。边的颜色和样式表示该调用所属的交易。



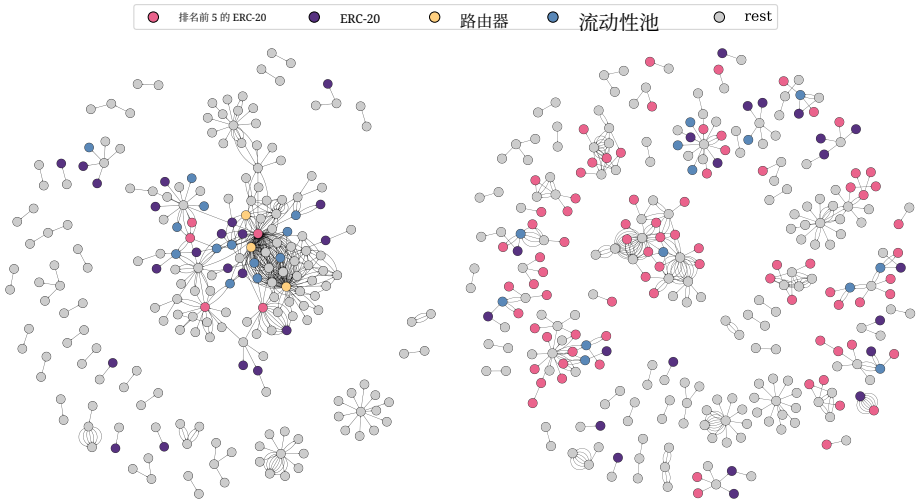
(b) 一组以太坊交易的基于交易的图表示，其中顶点为交易，边表示两笔交易之间的依赖关系。与同一地址交互的交易彼此依赖。

图 2：同一组五笔以太坊交易的两种图表示。边的颜色表示其在基于地址的图表示中所属的交易（参见图 2a）；基于交易的图表示则将相同的交易集作为顶点。

在区块 n、……、m 中出现的地址，即作为这些区块中某笔交易的一次（内部）调用的发送方或接收方的地址。E 是图的边集合，每条边 $e = (v, u)$ 表示区块 n、……、m 中合约 v 对合约 u 的一次调用。边 e, ω_e 的权重由相应内部调用所消耗的 *Gas* 量决定。

在定义 2 中，我们定义了基于交易的图，并在图 2b 中绘制了相应示例。请注意，基于地址的表示会导出基于交易的表示。在基于交易的图中，相邻交易无法安全地并行执行。因此，在图 2b 所示的示例中，交易 tx_4 不能与交易 tx_3 同时执行，因为它们都会与地址 D 交互（参见图 2a）。基于交易的表示中的团表明，团内的所有交易都必须按顺序执行。因此，交易 tx_1、tx_2 和 tx_3 无法并行化执行。下文中，我们主要讨论基于地址的表示，同时也会从基于交易的表示中获取独到见解，即计算图中最大的团，以探究并行化的极限。

定义 2（基于交易的图（Transaction-based Graph，TG$_{n,\ m}$））。对于区块 n，……，m、$n \leq m$，其基于交易的图表示为 *TG*$_{n,\ m}$（V、E、$\{\omega_v\}_{v \in V}$）。其中，V 是图的顶点集合，每个 $v \in V$ 都是一笔交易，而 V 是出现在区块 n，……，m中的所有交易组成的集合。顶点 v、ω_v的权重是交易 v所消耗的 *Gas* 数量。E 是图的边集合，每条边 $e = \{v, u\}$ 表示区块n，……，m中交易 v 与 u 之间的依赖关系。当两笔交易与同一地址交互时，即该地址分别作为两笔交易中某次（内部）调用的发送方或接收方出现时，便会产生依赖关系。



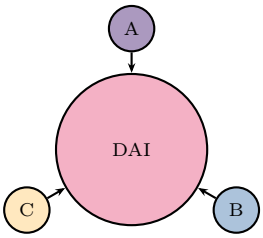
(a) 原始基于地址的图　　(b) 解耦的基于地址的图

图 3：区块 15,348,042（于 2022年8月15日挖出）的基于地址的可视化。图 3a 展示原始图，图 3b 则展示该图的解耦版本（参见第 6.1 节）。在这两幅图中，我们重点标出了若干核心 DeFi 智能合约，即 ERC-20 代币、去中心化交易所路由器和 DEX 流动性池。请注意，解耦图的最大连通分量明显小于原始图的最大连通分量。

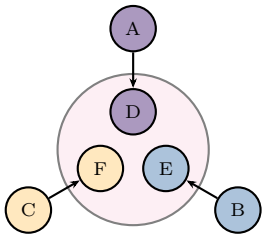
在图 3a 中，我们展示了区块 15,348,042 的地址交互情况（该区块于 2022年8月15日挖出），并突出显示了一些核心 DeFi 智能合约。粉色顶点是按转账数量计算排名前五的 ERC-20代币（WETH、USDC、USDT、DAI 和 LINK），紫色顶点则是出现在区块 15,348,042 中的其余 ERC-20 代币地址。我们用黄色突出显示以下去中心化交易所路由器：Uniswap V2、Uniswap V3、SushiSwap 和 1inch；并用蓝色标记 Uniswap V2、Uniswap V3、SushiSwap 和 Curve 流动性池。请注意，该区块的大部分内容构成了一个单一的连通分量，且绝大多数带标签的 DeFi 合约都属于这一连通分量。其中，一些合约（主要是排名前五的 ERC-20代币和去中心化交易所路由器）位于该连通分量的中心位置，因此对区块中不同交易之间的依赖关系作出了很大贡献。

6.1 解耦交易图表示

自去中心化金融兴起以来，在大多数区块中，这些去中心化金融合约在基于地址的交易图中始终保持较高的度。观察到这一点后，我们发现，这些智能合约引入的许多依赖关系并非必不可少，而这些合约是 DeFi 生态系统的核心组成部分。尤其是 ERC-20 代币合约和去中心化交易所路由器在智能合约层引入的明显依赖关系，在存储键层并不会显现。正如下文所述，验证者相对容易识别这两类非必要依赖关系。



(a) 基于地址的 ERC-20 转账图



(b) 解耦的基于地址的 ERC-20 转账图

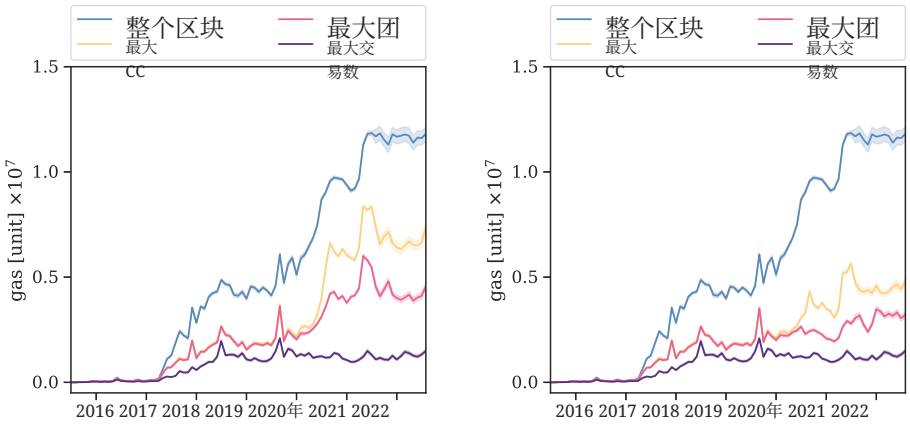
图 4：以下钱包之间三笔 DAI（一种 ERC-20 代币）转账的基于地址的图表示：从地址 A 到地址 D、从地址 B 到地址 E，以及从地址 C 到地址 F。图 4a 展示了实际的基于地址的表示。可以看到，这三笔转账似乎彼此依赖。图 4b 展示了我们如何解耦该图以避免这种依赖关系。

例如，考虑以下地址之间的三笔 DAI 转账。DAI 是一种 ERC-20 代币：地址 A 向地址 D转账，地址 B 向地址 E转账，以及地址 C 向地址 F转账。由于 DAI 并非以太坊区块链的原生货币，我们只能观察到 DAI 发送方对 DAI 智能合约的调用；该合约负责跟踪同质化 DAI 代币（参见图 4a）。因此，这三笔转账看起来彼此依赖，而三笔等价的 ETH 转账则不会如此。为此，我们对图 4b 中的交易图表示进行了解耦。我们不再假设交易的发送方调用代币合约，而是假设其调用 DAI 智能合约中接收方的内存位置。除了对 ERC-20 转账函数进行这些调整外，我们还对以下 ERC-20 合约函数作了相应调整：balanceOf、transferFrom、approve 和 allowance。请注意，我们只对排名前五的 ERC-20 代币进行这种解耦：WETH、USDC、USDT、DAI 和 LINK。这五种 ERC-20 代币的转账数量最多，合计占所有 ERC-20 转账的 34%。此外，我们选择仅限于这少数几种 ERC-20 代币，以展示：(1) 它们对区块内交易连通性的影响；(2) 确保这些代币不会出现意料之外的行为，例如向第三方转移一定比例的代币 [14]。

去中心化交易所路由器也参与了许多交易，因此会增加交易图的连通性。由于去中心化交易所路由器本身是无状态的，只代表用户调用指定的流动性池，因此智能合约层的依赖关系并非必要。因此，我们从交易图中移除路由器。具体而言，我们将所有路由器边重新连接到相应交易的发送方。在后续分析中，我们会对以下去中心化交易所的路由器进行这种解耦：Uniswap V2、Uniswap V3、SushiSwap 和 1inch。

7 可并行化性

在接下来的以太坊交易图分析中，我们量化了有限的并行化潜力。在第7.1节中，我们讨论了其随时间的演变——去中心化金融和非同质化代币的影响由此得以巩固——以及



(a) 原始交易数据 (b) 解耦交易数据

图5：我们绘制了以下各项所消耗的 Gas：(1) 整个区块，(2) 区块中最重的连通分量（CC），(3) 区块中最重的团，以及 (4) 区块中最重的交易。图5a分析原始交易数据，图5b分析解耦交易数据。请注意，我们绘制了月平均值，并附上从每天随机抽取的 65 个区块计算得出的 95%置信区间。

第7.2节中以太坊主网上并行化潜力的当前状况。我们将区块的可并行化性定义为可实现的最高加速比（区块使用的 Gas 总量除以某种调度的顺序 Gas）。在分析中，我们考察具体的调度以及图指标；根据我们对冲突的定义，这些指标为可并行化性提供上界和下界。

7.1 随时间变化的可并行化性

接下来，我们通过考察交易图的连通性，分析以太坊主网上的并行化潜力。我们在整个区块链历史中，每天随机抽取 65 个区块，直至 2022年8月31日的最后一个区块，从而观察随时间变化的趋势。

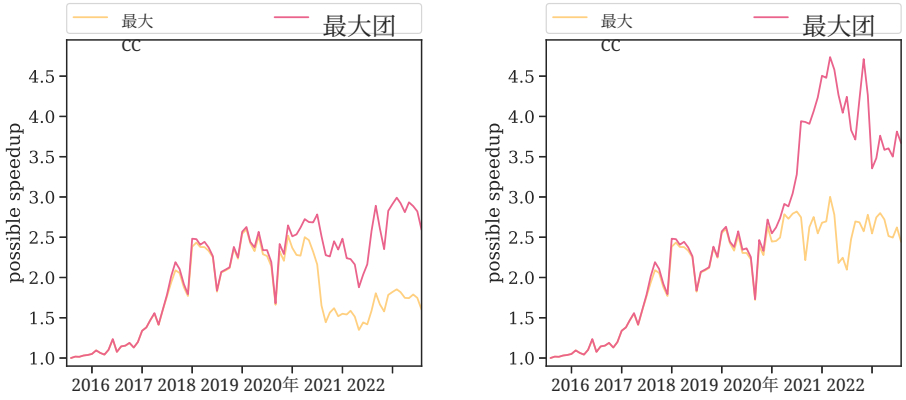
从 Gas 使用量随时间变化的趋势中，可以清楚地看到去中心化金融和NFT市场的普及情况；Gas 使用量是执行时间的代理指标。在图5a中，我们用蓝色绘制每个区块的 Gas 用量。请注意，随着 2020年去中心化金融的兴起，Gas 用量急剧增加。尽管这些新应用并未使交易数量达到前所未有的水平，但由于交易复杂性不断上升，它们导致每个区块的 Gas 用量飙升。

为了更深入地了解这些日益繁重的区块的可并行化性，我们还在图5a中绘出了最重连通分量的大小（黄色）和最重团的大小（粉色）。请注意，我们通过连通分量或团中交易所消耗的 Gas 总量来衡量其权重。因此，这些权重反映了执行其中交易所需的时间。

在基于地址的图表示和基于交易的图表示中，连通分量是等价的。在这两种视角下，最重连通分量的权重都为区块执行的并行化潜力提供了一个下界。任何尽可能并行执行交易的调度，也就是说，在每个时间步执行一个极大独立集中的交易，都不会超过顺序执行最重连通分量所需的时间。需要指出的是，要获取交易之间的依赖关系，必须能够访问静态提供的访问列表，或者先执行所有交易。不过，这项工作只需由验证者完成一次。区块执行完毕后，便可将并行调度提供给其他所有人用于验证。此外，我们利用基于交易的图表示来寻找最重团。在基于地址的图中，没有与之直接对应的概念。最重团为区块执行的并行化潜力规定了一个上界。假设交易原子执行，任何调度都必须顺序处理团中的所有交易。事实上，我们运行了一个简单的列表调度算法来生成调度。它会生成一个偏序，并始终允许并行执行一个极大独立集。我们发现，尽管该调度偶尔需要比最重团所需时间更长才能完成执行，但相对误差可以忽略不计（参见附录A）。因此，借助简单的调度，几乎可以达到并行化潜力的上界。请注意，由于我们是在智能合约层而非存储键层进行分析，交易图可能会高估依赖关系。

观察数据可知，目前最重连通分量占区块的一半以上（参见图5a）。此外，自 2020 年去中心化金融兴起以来，平均最重连通分量与平均最重团之间的差距有所扩大。这可能是由 DeFi 生态系统中不同协议和智能合约之间的交互所致。自去中心化金融兴起以来，最重连通分量中的交易往往会与热门 ERC-20 代币、DEX 流动性池和借贷协议交互。不过，通常只有与同一智能合约交互的交易才会形成一个团。值得注意的是，最大的团通常由与 WETH 交互的交易组成；WETH 是最受欢迎的 ERC-20 代币（按转账数量计）。

如前所述，与 WETH 交互的交易通常构成原始交易数据中最重的团。然而，正如我们在第 6.1 节所示，智能合约层视图中表面上的依赖关系，只不过是将 ERC-20 代币实现为智能合约而非原生代币所导致的结果。为了获得更准确的情况，我们进行了前文概述的解耦，并观察到自去中心化金融兴起以来，最重连通分量和团的规模均显著缩小（参见图5b）。事实上，从2020年7月1日到2022年8月31日，解耦使最重连通分量的规模平均缩小了1.78倍。此外，最重团的规模平均缩小了1.88倍。



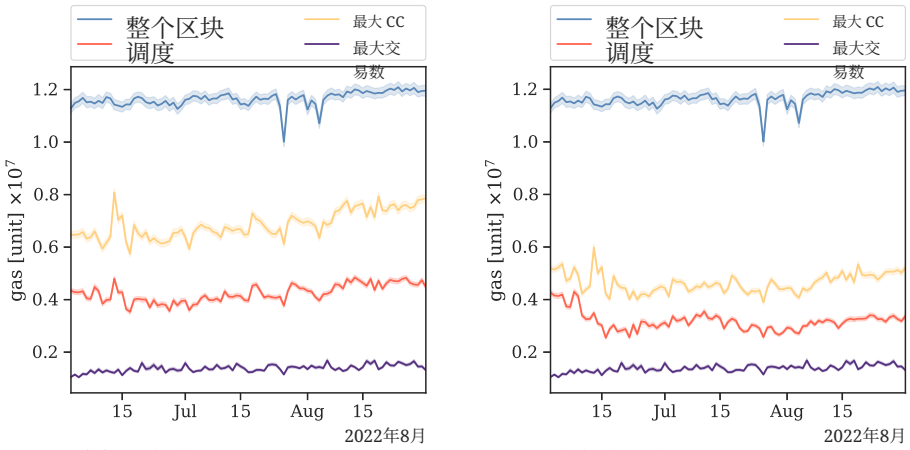
（a）原始交易数据　　（b）解耦交易数据

图6：我们通过对原始交易数据（参见图6a）和解耦交易数据（参见图6b）进行并行化，展示了可实现的执行加速比（按月汇总）。我们通过识别最重连通分量获得并行化潜力的下界，并通过最重团获得上界。请注意，最重连通分量和团均以 Gas 作为权重。

我们还在图5a和图5b中绘出了每个区块中最重交易的规模，以所用 Gas 衡量。[2] 区块中最重交易的规模，给出了并行化潜力的另一个更宽松的上界，该上界不考虑交易间的任何依赖关系。忽略所有依赖关系，就会自动排除任何非必要依赖关系。因此，这个更宽松的上界仅假设交易必须以原子方式执行。然而，我们发现，平均而言，最重交易在整个区块中占有相当大的比例——纵观全部历史数据，平均约为九分之一。因此，并行化不仅受不断增大的最重团规模限制，也同样受到单笔大型交易的制约。

我们在图 6 中绘出了可实现加速比的下界和（现实的）上界。图 6a 展示了原始交易数据对应的这些界限，图 6b 则展示了解耦交易数据对应的相同界限。与此前的观察类似，现实中可实现加速比的下界（由最重连通分量的大小决定）和上界（由最重团的大小决定）在 2020年 去中心化金融兴起之前彼此接近。此外，我们还观察到，所进行的解耦也仅从 2020年 起显现效果，因为它针对的是 DeFi 智能合约。从 2020年 起，可实现加速比的下界与上界之间的差距有所扩大。此外，在原始交易数据中，我们观察到，去中心化金融开始普及后，现实中可实现加速比的下界和上界都下降了（参见图 6a）。而在解耦

[2] 请注意，解耦不会影响最重交易的大小，也不会影响区块的总 Gas。



（a）原始交易数据 (b) 解耦交易数据

图 7：我们绘制了以下对象所消耗的 Gas：(1) 整个区块；(2) 区块中最重的连通分量（CC）；(3) 我们的朴素调度（顺序执行）；以及 (4) 区块中最重的交易。图 7a 分析原始交易数据，图 7b 分析解耦交易数据。请注意，我们绘制的是日平均值及其 95%置信区间。此外，我们使用调度的顺序 Gas 作为最重团规模的代理指标。

交易数据中，我们则注意到，引入去中心化金融后，可能实现的加速比下界并未下降，而是基本保持不变（参见图 6b）。更值得注意的是，对于解耦交易数据，引入去中心化金融后，现实中可实现加速比的上界甚至有所上升。我们推测，这源于同期交易数量不断增加（参见图 1a）。此外，这可能也受到以下因素的影响：在图表示中，大多数去中心化金融交易在给定数量的跳范围内彼此依赖，但不一定全部属于同一个团。

7.2 当前可并行化性的局限

为了更好地衡量当前可并行化性的局限，我们在先前分析的基础上，分析了 2022年6月1日 至 2022年8月31日 的所有区块，从而全面了解以太坊主网的当前状态。在图 7 中，我们绘制了以下对象所消耗的 Gas：(1) 整个区块；(2) 区块中最重的连通分量；(3) 我们的调度（顺序执行）；以及 (4) 区块中最重的交易。请注意，我们仅绘制调度所使用的顺序 Gas，而非最重团的 Gas，因为寻找最重团非常耗时。附录 A 中的分析表明，我们的调度几乎达到了相同的并行化潜力。

观察图 7 时，我们发现区块的日均大小和最重交易的日均大小都很少波动。我们只观察到两次整个区块平均大小骤降，幅度约为 10%，分别发生在 7 月底和 8 月初。观察每日

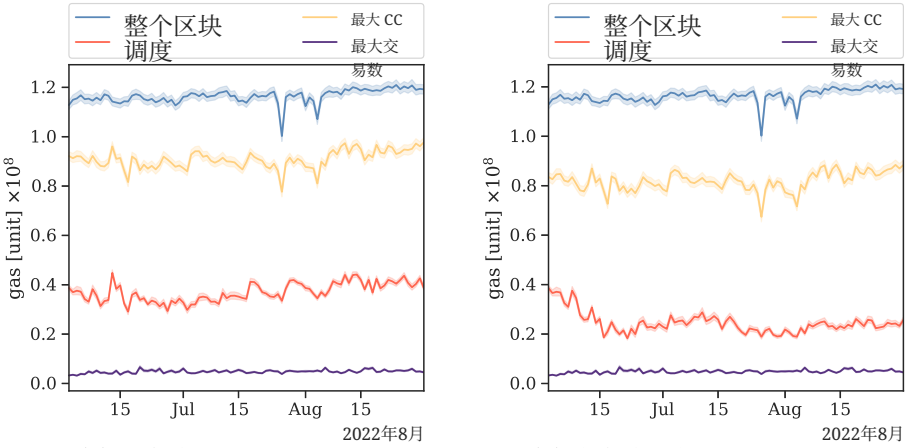
最重连通分量的平均大小，以及我们的调度在原始交易数据中使用的顺序 Gas 日平均量（参见图 7a），呈现出类似的情况。总体而言，两者的平均值分别约占区块大小的二分之一（连通分量）和三分之一（调度）。6 月 15 日左右，最重连通分量和调度使用的平均 Gas 都出现了一个峰值，而区块大小中并未观察到这一现象。由于市场预期 CPI 数据公布，且数据随后发布，以太币在此期间的每日价格变动非常剧烈（参见附录 C，图 12b） [41]。因此，Uniswap V3 等去中心化交易所上的以太币交易量迅速增加 [40]，我们推测，这导致最重连通分量和团相对于区块大小都有所增大。需要指出的是，尽管四张图中的日均值周围的 95%置信区间都很窄，但如附录 B 所示（参见图 11），四项图指标的数值波动都很大。例如，在我们观察到 Gas 使用量峰值的时段附近，最重交易的第99百分位数几乎达到区块大小的第99百分位数。这意味着有些区块中，单笔交易几乎占据了整个区块，因而这些区块几乎无法并行化。尽管如此，在近期的（原始）交易数据中，我们观察到，最重连通分量和最重团的 Gas 使用量日平均值（其中我们以调度作为代理指标）占区块的比例相对稳定。

再来看解耦交易数据（参见图 7b），我们注意到，最重连通分量所使用的 Gas 日平均值稳定下降了（下降至原来的 1.6 分之一），调度所使用的顺序 Gas 也稳定下降了（下降至原来的 1.5 分之一）。值得注意的是，六月初的降幅不如数据集其余部分显著。我们推测，这是同期以太币价格大幅下跌所致（参见图 12a），这可能引发了特殊的去中心化金融使用模式，使工作负载的关联性进一步增强。在数据集其余部分，解耦带来的降幅非常稳定，但可实现的加速比仍仅约为 4 倍（参见附录 B，图 10）。

最后，为了模拟更高的交易吞吐量，我们将连续十个区块作为一批，并考察相应交易图的连通性（参见图 8）。即使采用这种（相当夸张的）模拟区块大小增幅，我们调度的顺序 Gas 仍按比例增加，因此不会改变现实中可实现加速比的上界。另一方面，下界（由最重连通分量表示）则变得宽松得多。这表明，合并区块时，所有区块中最大的团会合并为一个，而连通分量甚至会在同一区块内部合并。这与我们的分析一致：最重团总是由相同的少数几个合约所诱导——这表明增大区块大小并不会提升并发潜力。

8 讨论与结论

我们的研究对以太坊主网工作负载的可并行化性进行了量化。



（a）原始交易数据　　（b）解耦交易数据

图8：我们绘制了以下对象所使用的 Gas：（1）连续的 10 个区块；（2）它们的最重连通分量（CC）；（3）我们的朴素调度（顺序执行）；以及（4）它们的最重交易。图 8a 分析原始交易数据，图 8b 分析解耦交易数据。请注意，我们同时绘制了日平均值及 95%置信区间。

我们发现，目前并发性水平非常有限。因此，仅仅致力于寻找充分利用现有并发潜力的方法还不够。相反，我们认为，部分重点必须转向确保工作负载本身具备可并行化性。具体而言，我们认为必须着力于以下三个方面，才能使现有并发机制实现更高的加速比。

调查依赖关系。 正如我们所述，去中心化金融的一些核心智能合约似乎出现在许多交易中。因此，我们认为，必须以更细粒度的方式调查交易依赖关系，例如细化到存储键级别。此外，还可以重新设计智能合约，以避免交易图中不必要的依赖关系。

激励“简单”交易。 目前，一个区块中的最重交易约占平均区块大小的十分之一。因此，这些单独的交易限制了并行化潜力。为此，我们认为区块链应当抑制此类频繁的高负载交易，转而鼓励简单交易。一种可行的方法是对计算收取超线性费用。

提高依赖关系的可预测性。 访问列表的使用率极低，这表明目前交易发送方无法提供其交易将访问的地址和存储键。提高依赖关系的可预测性可以解决这一问题，并有助于在执行期间实现更高程度的并行化。

只有当以太坊主网上的工作负载真正可并行化时，加速比才足以使以太坊基金会所述的每秒 100,000 笔交易 [9] 成为现实。

参考文献

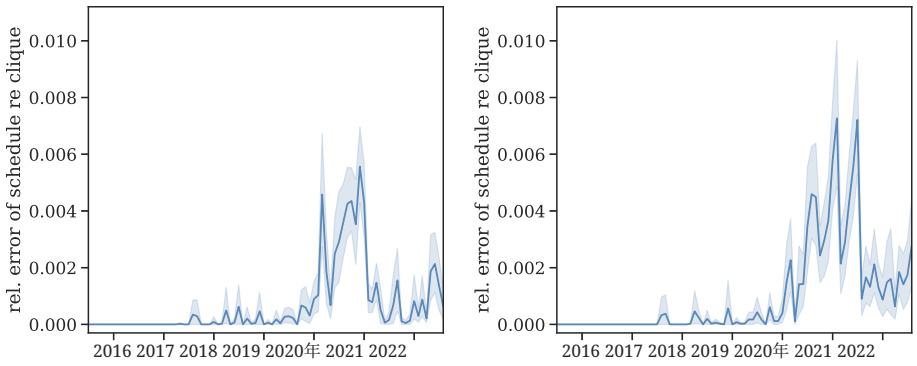
1. Amiri, M.J., Agrawal, D., El Abbadi, A.: Parblockchain：利用许可型区块链系统中的交易并行性。载于：2019 IEEE 第39届国际分布式计算系统会议（ICDCS），第1337–1347页。IEEE（2019） 2. Anjana, P.S., Kumari, S., Peri, S., Rathor, S., Somani, A.: Optsmart：一种节省空间的智能合约乐观并发执行方法。《分布式与并行数据库》，第1–53页（2022） 3. Avarikioti, G., Kokoris-Kogias, E., Wattenhofer, R.: 分而治之，扩展规模：分布式账本分片协议的形式化。arXiv 预印本 arXiv:1910.10434（2019） 4. Bai, Q., Zhang, C., Xu, Y., Chen, X., Wang, X.: 以太坊的演变：时序图视角。arXiv 预印本 arXiv:2001.05251（2020） 5. Berg, J.A., Fritsch, R., Heimbach, L., Wattenhofer, R.: Uniswap 和 SushiSwap 中市场低效的实证研究。载于：第二届去中心化金融（DeFi）研讨会，格林纳达（2022年2月） 6. Buterin, V., Conner, E., Dudley, R., Slipper, M., Norden, I., Bakhta, A.: EIP-1559。https://eips.ethereum.org/EIPS/eip-1559（2022） 7. Buterin, V., Swende, M.: EIP-2929：提高状态访问操作码的 Gas 成本。https://eips.ethereum.org/EIPS/eip-2929（2022） 8. Buterin, V., Swende, M.: EIP-2930：可选访问列表。https://eips.ethereum.org/EIPS/eip-2930（2022） 9. Cavicchioli, M.: 以太坊将达到每秒 100,000 笔交易。
h
t
t
ps://en.cryptono
mist.ch/2022/07/22/ethereum-reach-100000-transactions-second/（2022） 10. Chen, T., Li, Z., Zhang, Y., Luo, X., Chen, A., Yang, K., Hu, B., Zhu, T., Deng, S., Hu, T., 等：Dataether：以太坊数据探索框架。载于：2019 IEEE 第39届国际分布式计算系统会议（ICDCS），第1369–1380页。IEEE（2019） 11. Chen, Y., Guo, Z., Li, R., Chen, S., Zhou, L., Zhou, Y., Zhang, X.: Forerunner：面向以太坊的基于约束的推测性交易执行。载于：ACM SIGOPS 第28届操作系统原理研讨会论文集，第570–587页（2021） 12. Dang, H., Dinh, T.T.A., Loghin, D., Chang, E.C., Lin, Q., Ooi, B.C.: 通过分片实现区块链系统扩展。载于：2019年国际数据管理会议论文集，第123–140页（2019） 13. Decker, C., Wattenhofer, R.: 基于比特币双向微支付通道的快速可扩展支付网络。载于：第17届分布式系统稳定性、安全性与保障国际研讨会（SSS），加拿大埃德蒙顿（2015年8月） 14. DeFi Cartel：Salmonella。https://github.com/Defi-Cartel/salmonella（2022） 15. Dickerson, T., Gazzillo, P., Herlihy, M., Koskinen, E.: 为智能合约引入并发性。载于：ACM 分布式计算原理研讨会论文集，第303–312页（2017） 16. 以太坊基金会：ERC-20 代币标准。https://ethereum.org/en/developers/docs/standards/tokens/erc-20/（2022） 17. 以太坊基金会：Gas 与费用。https://ethereum.org/en/developers/docs/gas（2022） 18. 以太坊基金会：合并升级。https://ethereum.org/en/upgrades/merge/（2022）

19. 以太坊基金会：分片。https://ethereum.org/en/upgrades/ sharding/ (2022) 20. Ferretti, S., D'Angelo, G.：《以太坊区块链结构：复杂网络理论视角》。《并发与计算：实践与经验》**32**(12), e5493 (2020) 21. Gelashvili, R., Spiegelman, A., Xiang, Z., Danezis, G., Li, Z., Xia, Y., Zhou, R., Malkhi, D.：Block-stm：通过将排序难题转化为性能优势来扩展区块链执行。arXiv 预印本 arXiv:2203.06871 (2022) 22. Guo, D., Dong, J., Wang, K.：以太坊交易关系的图结构与统计特性。《信息科学》**492**, 58–71 (2019) 23. Han, R., Yu, J., Zhang, R.：分析并改进分片区块链的分片分配协议。密码学电子印本档案 (2020) 24. He, N., Su, W., Yu, Z., Liu, X., Zhao, F., Wang, H., Luo, X., Tyson, G., Wu, L., Guo, Y.：理解区块链生态系统的演变：对比特币、以太坊和 EOSIO 的纵向测量研究。arXiv 预印本 arXiv:2110.07534 (2021) 25. Kalodner, H., Goldfeder, S., Chen, X., Weinberg, S.M., Felten, E.W.：Arbitrum：可扩展的私有智能合约。载于：第27届 USENIX 安全研讨会（USENIX Security 18），第 1353–1370 页 (2018) 26. Kondor, D., Pósfai, M., Csabai, I., Vattay, G.：富者是否愈富？比特币交易网络的实证分析。《PLOS ONE》**9**(2), e86197 (2014) 27. ledgerwatch：Erigon。https://github.com/ledgerwatch/erigon (2022) 28. Lin, D., Chen, J., Wu, J., Zheng, Z.：以太坊交易关系的演变：通过微观模式理解全球驱动因素。IEEE计算社会系统汇刊 **9**(2), 559–570 (2021) 29. Lin, D., Wu, J., Yuan, Q., Zheng, Z.：通过复杂网络方法建模并理解以太坊交易记录。IEEE电路与系统 II：快报汇刊**67**(11), 2737–2741 (2020) 30. Luu, L., Narayanan, V., Zheng, C., Baweja, K., Gilbert, S., Saxena, P.：开放区块链的安全分片协议。载于：2016年 ACM SIGSAC计算机与通信安全会议论文集，第 17–30 页 (2016) 31. Motamed, A.P., Bahrak, B.：加密货币交易图的定量分析。《应用网络科学》**4**(1), 1–21 (2019) 32. Murgia, M., Galletta, L., Bartoletti, M.：区块链交易并行性理论。《计算机科学中的逻辑方法》**17** (2021) 33. Nakamoto, S.：比特币：一种点对点电子现金系统。《去中心化商业评论》，第 21260 页 (2008) 34. Optimism基金会：Optimism。https://www.optimism.io (2022) 35. Pîrlea, G., Kumar, A., Sergey, I.：结合所有权与可交换性分析的实用智能合约分片。载于：第42届 ACM SIGPLAN 编程语言设计与实现国际会议论文集，第 1327–1341 页 (2021) 36. Poon, J., Dryja：比特币闪电网络：可扩展的链下即时支付 (2016) 37. Ron, D., Shamir, A.：比特币完整交易图的定量分析。载于：国际金融密码学与数据安全会议，第 6–24 页。Springer (2013) 38. Saraph, V., Herlihy, M.：以太坊智能合约中推测性并发的实证研究。arXiv 预印本 arXiv:1901.01376 (2019)

39. Sergey, I., Hobor, A.：智能合约的并发视角。载于：国际金融密码学与数据安全会议，第 478–493 页。Springer (2017) 40. Uniswap Labs：以太币。https://info.uniswap.org/#/tokens/0xc02aaa39b223fe8d0a0e5c4f27ead9083c756cc2 (2022) 41. 美国劳工统计局：消费者价格指数发布时间表。https://www.bls.gov/schedule/news_release/cpi.htm (2022) 42. Wang, G., Shi, Z.J., Nixon, M., Han, S.：SoK：区块链分片。载于：第一届 ACM金融科技进展会议论文集，第 41–61 页 (2019) 43. Wood, G., 等：以太坊：安全的去中心化通用交易账本。《以太坊项目黄皮书》**151**(2014), 1–32 (2014) 44. Xie, Y., Jin, J., Zhang, J., Yu, S., Xuan, Q.：用于以太坊交易追踪的时序金额快照多重图。载于：区块链与可信系统国际会议，第 133–146 页。Springer (2021) 45. Xie, Y., Zhou, J., Wang, J., Zhang, J., Sheng, Y., Wu, J., Xuan, Q.：通过网络方法理解以太坊交易。载于：《图数据挖掘》，第 155–176 页。Springer (2021) 46. Zamani, M., Movahedi, M., Raykova, M.：Rapidchain：通过全分片扩展区块链。载于：2018年 ACM SIGSAC计算机与通信安全会议论文集，第 931–948 页 (2018) 47. Zanelatto Gavião Mascarenhas, J., Ziviani, A., Wehmuth, K., Vieira, A.B.：基于以太坊的加密货币交易动态研究。《复杂网络杂志》**8**(4), cnaa042 (2020) 48. Zhang, A., Zhang, K.：使用多版本排序实现智能合约并发。载于：亚太网络会议（APWeb）与网络时代信息管理（WAIM）网络与大数据联合国际会议，第 425–439 页。Springer (2018) 49. Zhao, L., Sen Gupta, S., Khan, A., Luo, R.：以太坊区块链网络的时序分析。载于：2021年网络会议论文集，第 2258–2269 页 (2021)

调度相对误差

我们在图 9 中比较了朴素调度所使用的顺序 Gas 与执行最重团所消耗的 Gas。请注意，在整个区块链历史期间，日均相对误差都非常小，不到 1%。从去中心化金融兴起到 2021 年年中，无论是原始交易数据还是解耦交易数据，相对误差都达到峰值。自 2021 年年中起，相对误差又进一步减小。我们还指出，对于解耦交易数据，我们的调度相对误差略高于原始数据。尽管如此，我们的朴素调度所消耗的顺序 Gas 能够准确近似执行最重团所消耗的 Gas；对于任何以地址为粒度处理冲突的调度而言，最重团都是最优的。

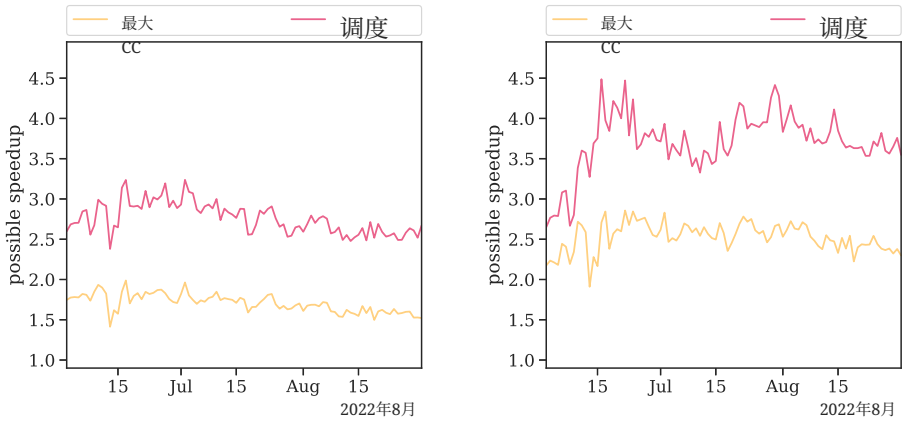


(a) 原始交易数据　　(b) 解耦交易数据

图9：我们展示了通过朴素调度算法得到的调度的顺序 Gas 与最重团之间的相对误差。对于任何在地址级别运行的并行调度，最重团都是可达到的最佳结果。我们绘制了原始交易数据（参见图9a）和解耦交易数据（参见图9b）中每天随机抽取的65个区块的日平均相对误差及95%置信区间。

B 当前可并行化性的局限

我们在图 10a 中绘制了原始交易数据的整个区块执行所用 Gas 的可实现执行加速比界限（按天取平均值），并在图 10b 中绘制了解耦交易数据的相应界限。按所用 Gas 计算，最重连通分量构成下界；而我们朴素调度算法所需的顺序 Gas（正如我们所见，它与最优调度十分接近）则可作为上界的代理指标



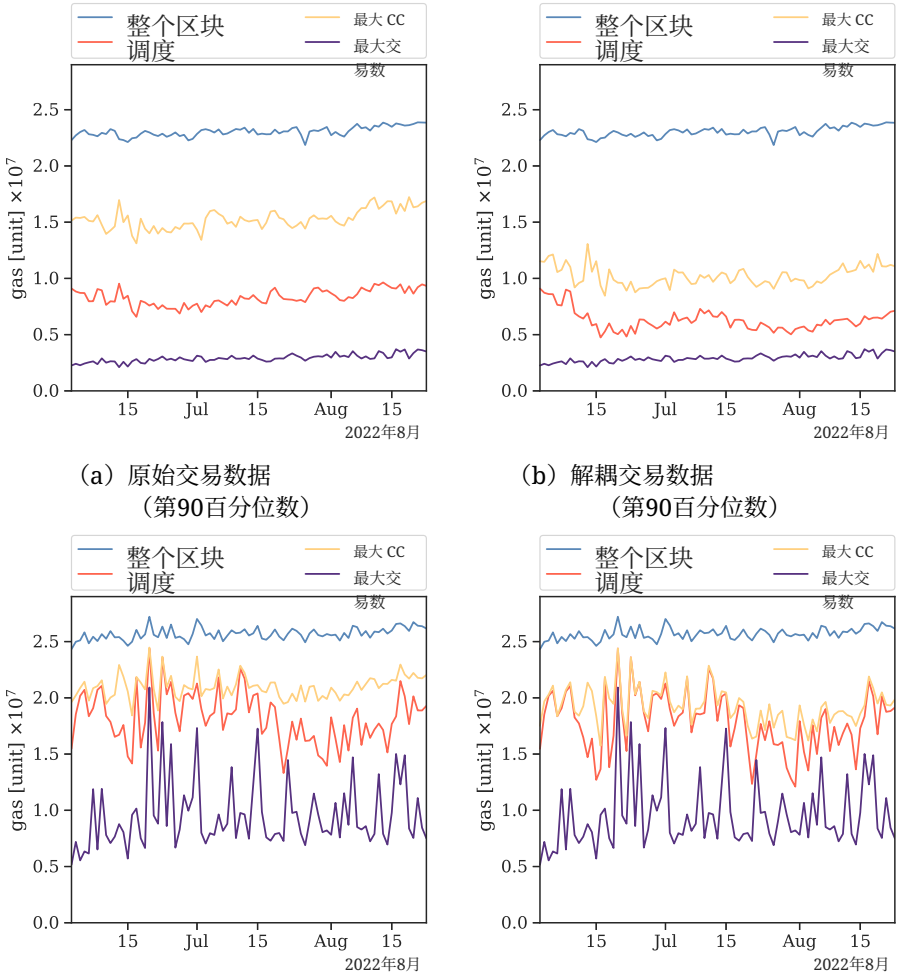
(a) 原始交易数据 (b) 解耦交易数据

图10：我们展示了原始交易数据（参见图10a）和解耦交易数据（参见图10b）通过并行化实现的执行加速比（日聚合值）。通过识别最重连通分量，得到并行化潜力的下界；通过最重团，得到其上界。请注意，最重连通分量和最重团的权重均按 Gas 计算。

。我们发现，即使纳入解耦处理，加速比也未超过 4.5 倍；尽管我们考察的只是理论界限，并未对并行度施加限制，例如未设定最大线程数。

图11a展示了原始交易数据中以下各项所用 Gas 的第90百分位数：(1) 整个区块的顺序执行，(2) 最重连通分量，(3) 我们的朴素调度，以及(4) 最重交易；图11b则展示了解耦交易数据中的相应数据。请注意，与日均值（参见图7a）相比，这四项指标的 Gas 日第90百分位数均约为其两倍。因此，10%的区块大小是平均区块的两倍。请注意，EIP-1559允许区块大小最多达到目标区块大小的两倍 [6]，而我们观察到相当大比例的区块大小非常接近目标区块大小的两倍。我们还观察到许多几乎为空的区块，其出现频率与超满区块相同；我们推测，这种空区块与超满区块交替出现的模式，是为了操纵 EIP-1559 引入的基础费用。与均值相比，最重连通分量和我们的朴素调度所用 Gas 的第90百分位数也出现了类似的增长，这表明在按 Gas 用量计处于第90百分位数的区块中，连通程度与平均区块相近。也就是说，区块中最重连通分量和团所占的比例基本保持不变。请注意，原始交易数据和解耦交易数据均是如此。

如果转而关注 Gas 使用量的第99百分位数（参见图 11c 和 11d），情况则呈现出不同的面貌。我们不再只是观察到成比例的



（a）原始交易数据
（第90百分位数）

（b）解耦交易数据
（第90百分位数）

（c）原始交易数据
（第99百分位数）

（d）解耦交易数据
（第99百分位数）

图 11：我们绘制了以下项目所使用 Gas 的每日第90和第99百分位数：（1）整个区块；（2）区块中最重的连通分量（CC）；（3）我们的朴素并行调度（顺序 Gas）；以及（4）区块中最重的交易。图 11a 和图 11c 分析原始交易数据，图 11b 和图 11d 分析解耦交易数据。请注意，我们将调度所使用的顺序 Gas 作为最重团大小的代理指标。

Gas 使用量增加，而是观察到截然不同的模式。只有整个区块的 Gas 使用量第99百分位数看起来保持稳定，因为它受到 EIP-1559 的限制。最重连通分量和最重团的第99百分位数都显著大于平均区块大小（参见图 7）。对于原始交易和解耦后的交易，这一点均成立

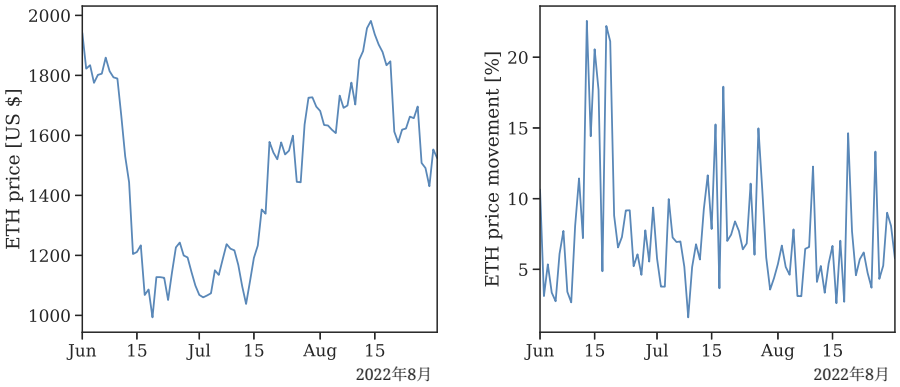
数据。在第99百分位数下，我们的解耦对最重团的影响很小，但仍会略微减小最大连通分量的规模。我们推测，这是由于这些特殊情况下最重团中的交易性质不同所致。此外，最大交易的第99百分位数与平均区块大小相当（参见图7），约为区块大小第99百分位数的一半。因此，最大交易占区块的比例相当大，从而极大地限制了这些区块的并行化潜力。

C 近期以太币价格数据

在图12中，我们绘制了2022年6月1日至2022年8月31日期间以太币的每日价格和每日价格变动。图12a展示了以太币的每日开盘价（p_{open}）和每日收盘价（p_{close}）。我们在图12b中绘制的每日价格变动，是衡量每日价格波动性的指标，其定义为

$$\frac{p_{high} - p_{low}}{p_{low}},$$

其中，p_{high} 为当日以太币最高价，p_{low} 为当日以太币最低价 [5]。



(a) 以太币价格　　(b) 以太币价格变动

图12：我们绘制了2022年6月1日至2022年8月31日期间的以太币价格和每日价格变动。每日价格变动比较了当日最高价与最低价，因此是衡量价格波动性的指标。