香 港 大 學
THE UNIVERSITY OF HONG KONG

基于GPU加速同态加密和关联事务融合的高性能加密数据库

关荣鑫

发明背景-区块链技术

- 区块链是一种以区块 (block) 为基本单位的链状数据结构。
  - 每个区块包含一定数量的、全序的事务 (transaction)。
  - 每个区块包含前一个区块的哈希值，从而构成一个链状结构。
- 区块链的所有节点运行在一个去中心化的点对点 (P2P) 网络中。
  - 节点之间通过分布式共识算法就确认 (commit) 的区块的内容和顺序达成共识。
  - 每个节点本地未确认的区块链可能会有临时的分叉 (fork)，但是共识算法保证已确认区块的强一致性
- 区块链的特性：
  - 不可篡改性：哈希算法的单向性与区块链的链状结构保证了已确认的数据无法被轻易篡改。
  - 数据可靠性：每个节点存储全体数据，部分节点故障不会导致数据丢失。
  - 高可用性：部分节点故障时仍可以继续添加和确认新的区块。
  - 透明性：所有关联方随时可以检查链上数据未被篡改。

发明背景-区块链分类

区块链可分为非认证型 (Permissionless) 区块链和认证型 (Permissioned) 区块链。

- 非认证型区块链不设立节点准入机制，所有节点匿名并且可以随时加入网络。
  - 由于节点匿名，非认证型区块链必须依赖加密货币来约束节点遵守网络规范。但加密货币通常需要消耗节点的大量计算资源或资金投入。
  - 为提高安全性，增加恶意节点生成支链和孤立区块的难度，非认证型区块链会增加区块生成的难度 (例如比特币的平均区块生成间隔为 10 分钟)，这大大降低了系统的吞吐量。
  - 因此，非认证型区块链的高开销和低性能使其主要用于加密货币（例如比特币、以太坊等），而**不适合**用于通用的数据分享平台。
- 认证型区块链设立节点准入机制，所有节点身份公开可验证，因此无需引入加密货币机制。此外，智能合约赋能通用计算能力。这使得认证型区块链**适合**用作通用的数据计算和分享平台。

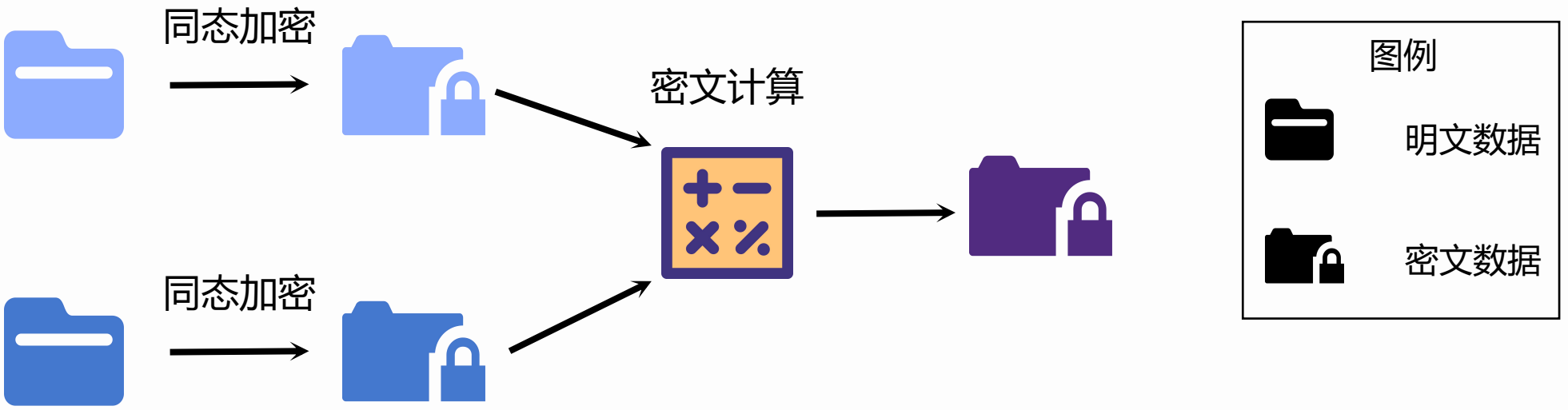
发明背景-现有区块链存在的数据隐私问题

现有区块链面临三大数据隐私问题：

- 隐私泄露：由于区块链的交易历史是公开可查的，因此如果链上数据以未加密明文形式存储，则面临数据隐私泄露风险。
- 隐私保护与计算性能的矛盾：为了实现隐私保护，区块链需要结合密码学技术，从而带来加密、解密与证明计算结果正确性的高昂开销，导致区块链计算性能恶化。
- 隐私保护与通用计算能力的矛盾：如果链上数据经密码学加密且以密文形式存储，则难以应用于具备通用计算能力的智能合约，导致隐私保护区块链的使用场景比较有限。

发明背景-同态加密技术

同态加密 (Homomorphic Encryption) 是一种基于特定代数结构的加密技术，它可以**直接对密文进行计算**，而无需先解密再基于明文进行计算。

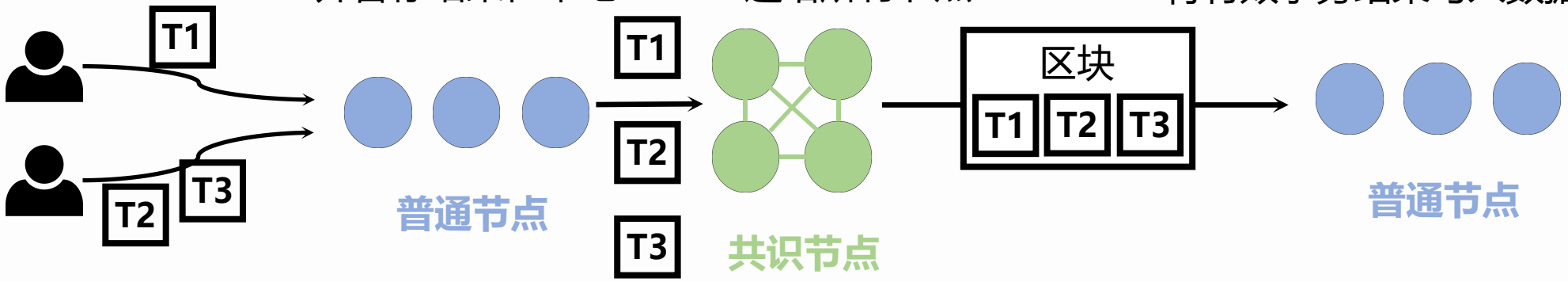


现有工作-“执行→共识→验证” 区块链工作流

① 客户提交事务到普通节点，节点执行事务并暂存结果在本地

② 共识节点将事务打包成区块，发送给所有节点

③ 普通节点按照区块中的事务顺序验证本地结果的有效性，并将有效事务结果写入数据库



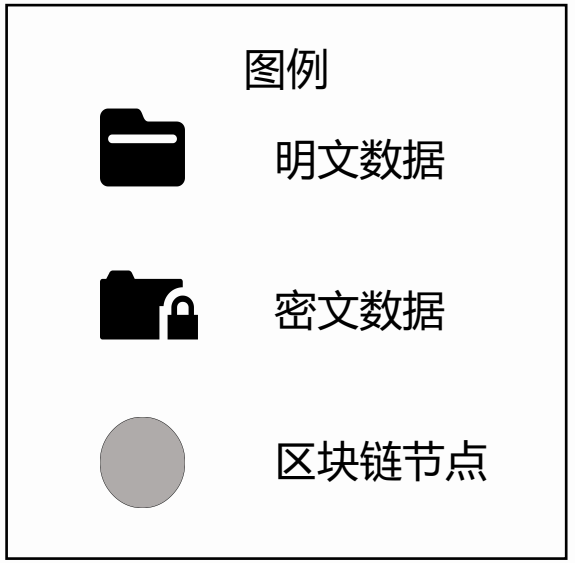
优点

- 高性能：并发度高、高吞吐量、低延迟
- 支持非确定 (non-deterministic) 的智能合约，可与含随机噪音的同态加密技术结合

现有工作-数据隐私保护区块链

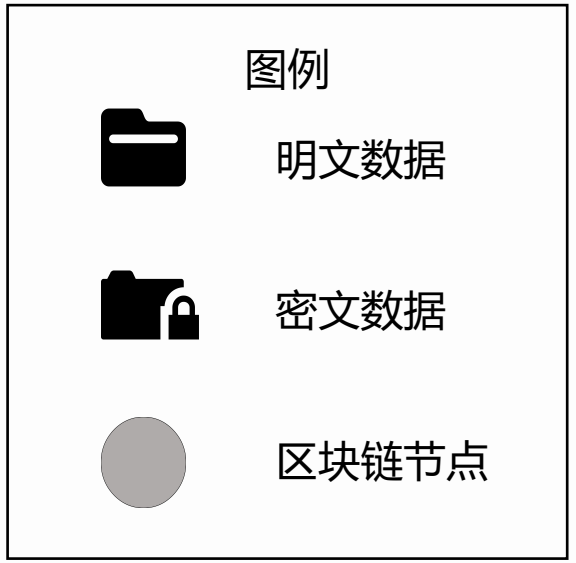
当前的数据隐私保护区块链可以分为两类：

1. 非全周期数据加密保护的区块链：进行链上计算时，先对链上密文数据进行解密，再基于明文计算，最终对结果进行再加密和存储。



现有工作-数据隐私保护区块链

2. 基于同态加密技术的全周期数据加密保护的区块链：进行链上计算时，直接基于密文数据进行计算和存储，无需解密和再加密。



现有工作-数据隐私保护与计算性能的矛盾分析

数据隐私保护区块链需要频繁进行加解密、同态密文计算、计算结果正确性证明（非交互式零知识证明）等开销高昂的密码学运算，显著降低了区块链的计算性能。

表 1 展示了常见数学运算的非密码学版本和 128-bit 安全级别的同态加密版本的最低时间开销。

表 2 展示了在 128-bit 安全级别下其他常见密码学计算的最低时间开销。

数学运算	非密码学版本（纳秒）	同态加密版本（纳秒）
58-bit 整数加法	1	1298
58-bit 整数乘法	1	3274000

表 1

数学运算	时间开销（微秒）
同态加密	1263
同态解密	276
非交互式零知识证明生成	50000
非交互式零知识证明验证	1000

表 2

现有工作-数据隐私保护与通用计算能力的矛盾分析

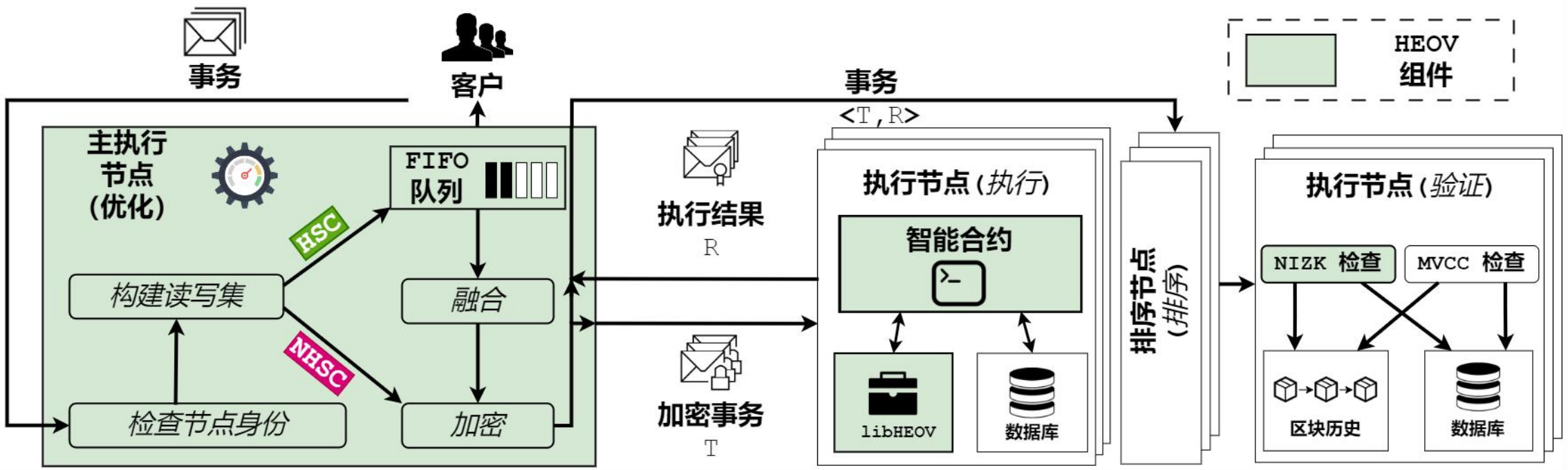
- 为了降低数据隐私保护的性能开销，部分现有工作采用部分同态加密 (Partially Homomorphic Encryption) 技术，然而这导致该部分工作只能支持有限种类的计算（例如仅支持密文加法，不支持密文乘法），使其不具备通用计算能力。
- 为了同时实现数据隐私保护和区块链智能合约的通用计算属性，部分现有工作朴素地采用了同时支持密文加法和乘法运算的全同态加密 (Fully Homomorphic Encryption) 技术，实现智能合约的图灵完备性。
  - 然而，全同态加密技术与数据隐私保护区块链的朴素结合导致极高的事务延迟 (百秒级别) 和内存占用 (百GB级别)，损害了数据隐私保护区块链的通用计算能力和实际应用场景。
  - 因此，需要将全同态加密技术与区块链工作流有机整合起来，方可同时实现数据隐私保护与通用计算能力。

本发明技术方案-主要思想

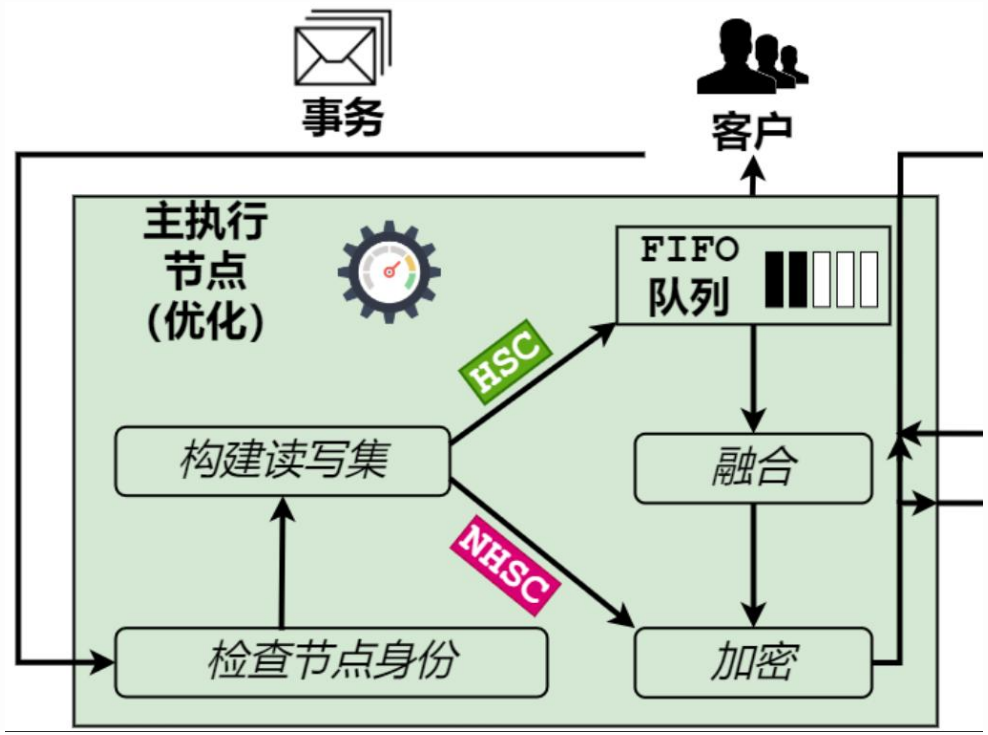
本发明在“执行 → 共识 → 验证”工作流的基础上，提出了 **区块链的全周期同态加密数据工作流** 与 **基于在线分析的关联事务融合技术**，实现了高性能的、保护数据全周期隐私的、具备通用计算能力的认证型区块链。

技术保护点1-全周期同态加密数据工作流

本发明利用“执行 → 共识 → 验证”工作流的背书策略 (Endorsement Policy) 特性，通过全同态加密和非交互式零知识证明的协同设计，设计了可实现全周期数据加密的区块链工作流。



技术保护点2-基于在线分析的关联事务融合技术



本发明技术方案-总结

与现有的区块链相比：

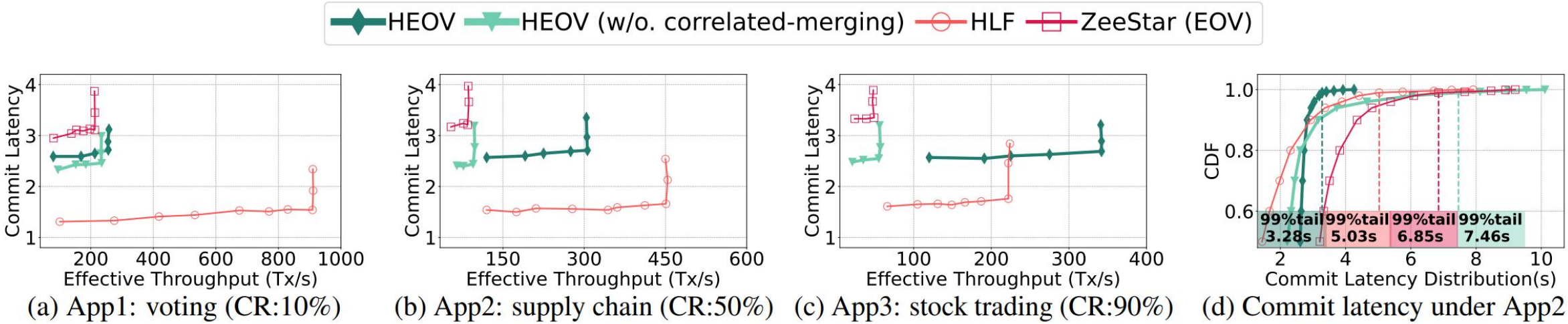
系统分类	数据隐私保护	直接密文计算	通用计算能力支持	性能
非加密区块链	✗	✗	✔	高
非全周期数据加密保护的区块链	✔	✗	✗	低
其他基于同态加密技术的全周期数据加密保护的区块链	✔	✔	✗	低
本发明HEOV	✔	✔	✔	高

实验结果-1/3

- 对比系统
  - HyperLedger Fabric [EuroSys 18]：当前主流的高性能开源认证型区块链
  - ZeeStar [S&P 22]：与本发明方向最贴近的数据隐私保护区块链
  - HEOV (无关联事务融合)：未采用关联事务融合技术的 HEOV 原型
- 实验环境
  - 20 台机器，各有 24 核 CPU, 64 GB RAM 和 40 Gbps 网络
- 目标问题
  - 与对比系统相比，本发明能否实现数据隐私保护通用智能合约的高性能执行?
  - 本发明能否抵御针对“执行 → 共识 → 验证”工作流的无重新提交攻击?

实验结果-2/3

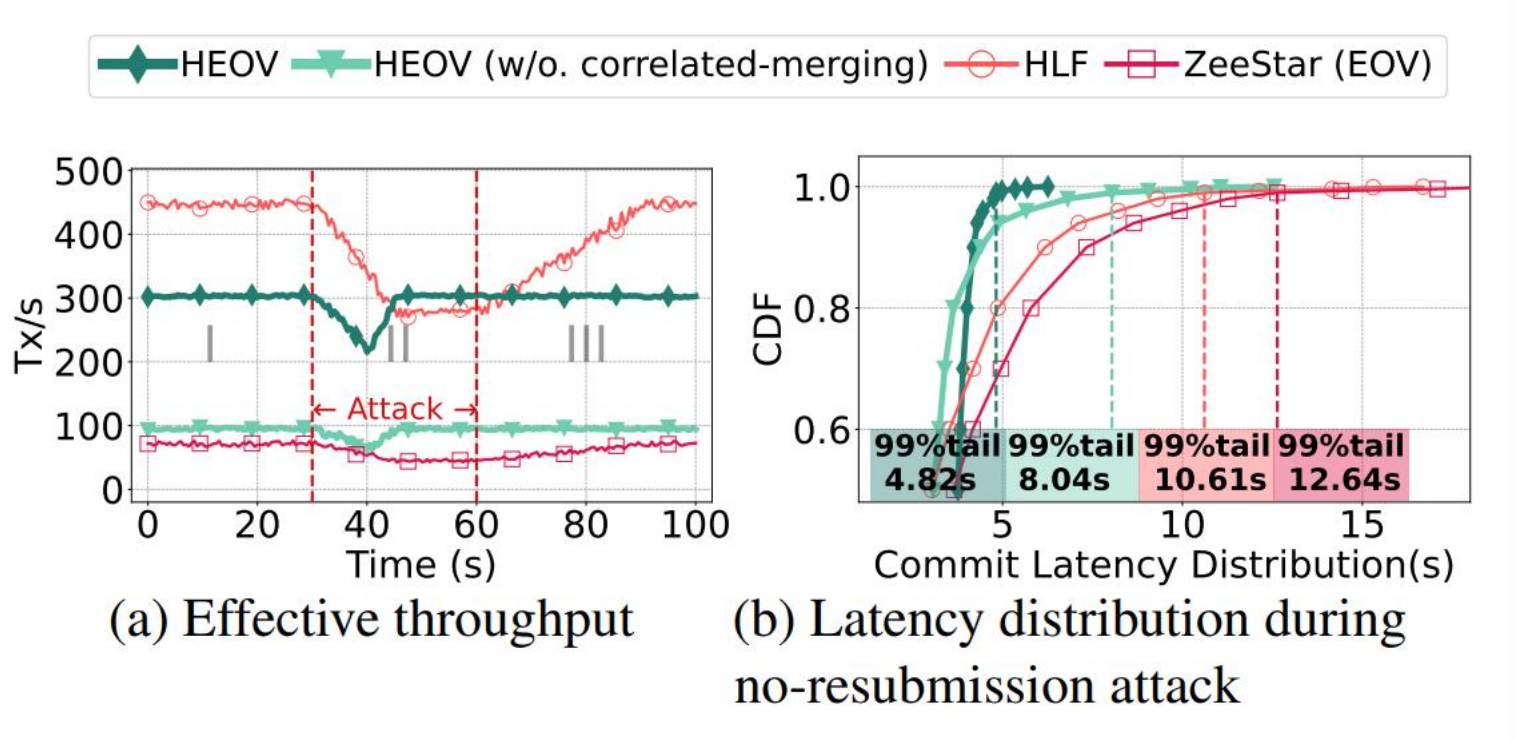
- 本实验对比了本发明与三个对比系统在无恶意攻击的良好网络下的性能
- 图 (a), (b), (c) 分别比较了在事务冲突率在 10%, 50% 和 90% 情况下的吞吐量和事务平均确认延迟
- 图 (d) 比较了在事务冲突率为 50 % 情况下事务确认延迟的连续分布



(a) App1: voting (CR:10%) (b) App2: supply chain (CR:50%) (c) App3: stock trading (CR:90%) (d) Commit latency under App2

实验结果-3/3

- 本实验对比了本发明与三个对比系统在无重新提交攻击 (No-Resubmission Attack) 下的性能
- 图 (a) 的竖直红色虚线划分了三个区域，从左至右分别比较了攻击前、攻击时、攻击后的各系统吞吐量
- 图 (b) 比较了在攻击发生时事务确认延迟的连续分布



(a) Effective throughput

(b) Latency distribution during no-resubmission attack

本发明的专利技术保护点

本发明是第一个提供全周期数据隐私保护、支持通用智能合约、高吞吐量低延迟的认证型区块链系统。本发明提供了两个专利技术保护点，协同实现了数据隐私保护和高吞吐量低延迟两个特性。

- 技术保护点1：全周期同态加密数据工作流
- 技术保护点2：基于在线分析的关联事务融合技术

技术保护点的侵权检测-1/2

针对某认证型区块链系统的具体侵权分析如下：

- 保护点 1：针对认证型区块链的全周期同态加密数据工作流。
    1. 第一步：作为一个客户加入该系统。
    2. 第二步：判断是否生成计算结果一致性非交互式零知识证明。
        1. 提交一个合法的事务给该系统，获取初步执行结果。
        2. 如果客户无需自行模拟智能合约得到结果并生成与其他节点执行结果的一致性证明，则该系统生成计算结果一致性非交互式零知识证明。
    3. 第三步：判断是否验证计算结果一致性非交互式零知识证明
        1. 统计从提交事务到确认事务上链的时间消耗。
        2. 将该时间消耗与先验知识进行比较，如果与验证计算结果一致性非交互式零知识证明的时间开销相近，则可以判断节点验证了计算结果一致性非交互式零知识证明。

即：若一个满足本发明目标的系统生成并验证计算结果一致性非交互式零知识证明，则该系统极有可能对本发明的**技术保护点 1** 造成侵权。

技术保护点的侵权检测-2/2

- 保护点 2：基于在线分析的关联事务融合技术。
    1. 第一步：作为一个客户加入该系统。
    2. 第二步：判断是否融合相关事务。
        1. 多次连续提交 N (N>1) 个相关事务给该系统。
        2. 如果需要生成远少于 N 个事务的计算结果一致性非交互式零知识证明，则可以判断该系统对相关事务进行融合。
    3. 第三步：判断是否不融合不相关事务
        1. 多次连续提交 N (N>1) 个不相关事务给该系统。
        2. 如果需要生成 N 个事务的计算结果一致性非交互式零知识证明，则可以判断该系统不对不相关事务进行融合。

即：若一个满足本发明目标的系统声称可减少智能合约计算量，且存在上述外显特征，则系统极有可能对本发明的**技术保护点 2** 造成侵权。