一种新型的半去中心化区块链架构*

杨 城[1,2] 罗旭斌[1,2] 段 江[1,2,3] 康 立[1,2] 颜 河[3]

[1](西南财经大学 经济信息工程学院 四川成都 611130)

[2](西南财经大学 中国区块链研究中心 四川成都 611130)

[3](成都恒图科技有限责任公司 四川成都 610041)

通讯作者: 罗旭斌, E-mail: xbluo@swufe.edu.cn

摘 要: 针对传统中心化的应用系统高效但不透明，而新兴去中心化的应用系统公开透明但共识机制低效的问题，本文提出了一种混搭架构的新型数据应用体系“半去中心化模式”。它将中心化的数据存储与去中心化的数据验证相结合，从而兼有中心化管理的高效性和分布式机制的透明性，其本质上是一种民主监督下的中心模式，兼顾了隐私与监督之间的平衡。该模式的核心思想是“类区块链存储”技术和“分布式集合查验”机制，二者共同维护体系的数据共识，通过三个合法性检验（身份合法性、存在合法性、数据唯一性）保证了数据的真实可靠性.

关键词: 半去中心化; 区块链; 分布式审查; Merkle 树; 集合验证

中图法分类号: TP309.2

A Novel Semi-decentralized Blockchain Architecture

YANG Cheng[1,2] LUO Xubin[1,2] DUAN Jiang[1,2,3] KANG Li[1,2] YAN He[3]

[1](School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu 611130, China)

[2](Blockchain Research Center, Southwestern University of Finance and Economics, Chengdu 611130, China)

[3](Chengdu Everimaging Ltd, 610041, China)

Abstract: Given the fact that the traditional centralized data application system is efficient but non-transparent, while the emerging decentralized data application system is transparent but inefficient in the consensus mechanism, this paper suggests a new trusted data application system in mixed architecture called “semi-centralized architecture”. Combining the centralized data storage mode and decentralized data validation method, it is both efficient and transparent. Essentially, the semi-centralized schema is a centralized approach based on distributed democratic supervision, taking into account the balance between privacy and supervision. The basic methods of the new schema are “similar blockchain storage” technology and “distributed validation by set” mechanism, which guarantee the authenticity and security of data in the whole system. And it ensures the reliability of the data through three legitimacy tests (identity legitimacy, existence legitimacy, and data uniqueness).

Key words: Semi-decentralized; Blockchain; Distributed Audit; Merkle Tree; Set Validation

1 引言

传统的数据应用系统，形式上无论是单中心系统还是分布式系统，数据的管理权限都控制在单个实体中，就管理主体而言实质都是中心化的应用模式（Centralized Architecture，CA），即数据的存储、维护和查询都由中心来统一实施，所有数据副本之间的一致性也由中心来统一协调。该模式的优势在于数据管理上的高效性，但存储的数据对外是一个“黑箱”，缺乏透明性与可监督性，客户只能被动接受，难以保证数据的真实性和正确性。社会信任本质上是一种公共认知[1]，公众对 CA 模式的信任主要源于对国家政府、明星企业以及社会名人的信任。尤其在网络环境下，其对象涉及整个应用系统服务链，公众对网络服务提供者

*基金项目: 教育部人文社科项目(17YJCZH210); 中央高校科研业务费专项资助项目(JBK120505)

Foundation item: MOE (Ministry of Education in China) Project of Humanities and Social Sciences (17YJCZH210); Fundamental Research Funds for the Central Universities (JBK120505).

的信任始终是公共信息服务的基础。但是，近年来因内鬼作案、系统漏洞或制度缺陷导致的用户数据遭篡改或泄露的事件时有发生，引发了民众对中心化应用模式的质疑，其公众信任问题面临巨大挑战。

新兴的区块链技术是一种加密的分布式记账技术，它由多方共同维护，以块链结构存储数据，并以密码学方式保证其不可篡改和不可伪造的去中心化的应用模式（Decentralized Archi，DA）[2-5]。在这种模式下，人人都是资源的提供方和获取方，没有绝对的中心，数据的共识性通过每个节点的完全冗余备份来实现。当某个节点被篡改时，它将显得“格格不入”，因而被排斥且不被承认。但 DA 模式在展示高度安全、透明的同时，其内在的低效数据共识机制也造成时间、空间和能耗上的极大浪费。此外，作为一个完全分布式的自治系统，该模式缺乏一个可靠的、公正的责任组织来协调各方利益和处理各种紧急事务。即使各利益方代表组成所谓的利益共同体，出于对自身利益最大化的考虑，在规则制定和事务协调上，该组织也难以秉持客观公正的态度有效的发挥作用。近年来的“THE DAO 事件”、“门头沟事件”和比特币的多次分叉，都集中反映出这种模式的潜在危机和内在缺陷[6][7]。

许多学者也都表达出类似的观点，对两种数据应用模式，尤其是后一种新兴模式的低效机制和可能引发的社会问题、安全问题表示出担忧。例如，Xiwei X.，Cesare P.等人认为，作为非中心化的区块链技术，虽然增加了安全性和透明性，却无法保证信息的私密性，不利于保护个人隐私[8]。Roman B.，Michel A.等人则对完全去中心化存在的矿池算力集中导致的 51%攻击危机、责任主体缺失导致的管理漏洞、低效共识机制导致的高时延和资源浪费等问题提出质疑[9]。此外，近年来区块链行业安全事故频出，业界专家和学者普遍认为，在安全性未得到完全保障的前提下，目前距离区块链应用大规模落地还有很长的路要走。但这些研究在指出问题与忧虑的同时，并没有提出一个可行的替代模式和解决方案。有鉴于此，本文将结合两种数据应用模式的优势和特点，设计一种混搭架构的新型数据应用模式，使其在保持安全高效、低成本运作的同时，兼备透明可监督的特点。

本文后续部分的结构如下：第二节整体性阐述该新型数据应用模式的设计理念和核心思想；第三节和第四节分别从“类区块链存储”和“分布式集合查验”这两个核心环节对新模式进行详细解读；第五节是新模式的效率分析；最后是全文的总结。

2 半去中心化模式

如前所述，中心化模式的高效便捷源于数据的集中管理和对中心的无条件信任，而去中心化模式的安全透明源于结合密码学技术的分布式存储。基于此本文提出了一种混搭架构的数据应用模式理念：对业务数据依旧采用集中存储，但在数据结构上结合区块链技术生成验证数据（业务数据的哈希摘要），以锁定业务数据，并将验证数据分散到所有客户端，从而达到分布式监督的效果。我们将这种模式称为“半去中心化”（Semi-Decentralized Archi，SDA），它的核心思想是“类区块链存储”和"分布式集合查验"。相对而言，前者是底层技术手段，后者是上层制度设计，二者共同维系整个体系的数据共识。

与区块链是比特币的底层技术类似，“类区块链存储”是 SDA 模式的技术基础，它着重体现在数据的存储结构上：数据记录按照某种用户可检索验证的顺序（如字母顺序、时间顺序等）依次存储在一个个区块内，所有区块按先后顺序相连，结成一种环环相扣的链式结构，这些结构化的区块被存储到数据库中，并以密码学技术保证数据的完整性、可追溯性和不可篡改性。差别在于，比特币的区块链是分布式存储的，拥有无数的完整副本，由大量矿工来共同维护；而“类区块链”则是由数据中心（简称“中心”）来集中存储和维护，可视为单矿工模式的区块链。

“分布式集合审查”是 SDA 模式的关键机制，它体现在用户对数据中心的民主监督上：一方面所有的数据服务，包括存储、管理、维护和查询等，都由中心集中高效的提供，另一方面每一个客户端都保存有所有区块的数据指纹和部分数据的哈希摘要。用户像传统 C/S 模式一样，通过手机、电脑向中心提请数据服务，同时自身也提供对中心的监督义务，即通过对中心数据的随机抽取和查验来防范和监督中心伪造或篡改数据，从而确保数据的安全可信。

图 1 展示了三种应用模式的对比关系。（a）CA 模式只有一份数据（或一个数据管理者），所有用户无条件信任中心。（b）DA 模式拥有无数份完全数据副本，体系依靠高冗余、高能耗和高网络带宽来维系整体

的数据共识。（c）SDA 模式由一份完全数据与无数份摘要数据片段共同构成。其中，数据中心（Center Database，CDB）拥有完整的业务数据 Data 和每一条记录的哈希摘要H(Data)；每个客户端的本地数据库（Local Database，LDB）仅存放摘要数据的一个子集H′(Data)。

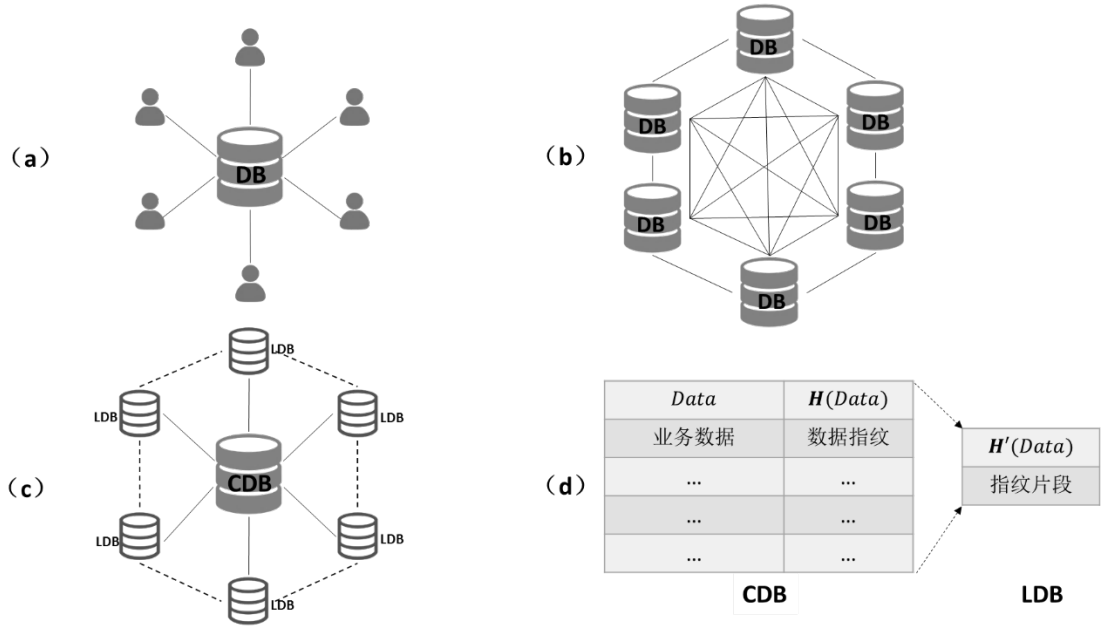

Fig.1 The comparison of three application system: (a) the centralized architecture (CA); (b) the full decentralized architecture (FDA) with full data replica; (c) the proposed semi-decentralized architecture (SDA); (d) the data structure and distribution of SDA.

图 1 三种应用体系的对比图，其中（a）中心化模式 CA；（b）去中心化模式 DA；（c）半去中心化模式 SDA；（d）SDA 模式的数据构成

综合而言，“半去中心化”本质上是一种民主监督下的中心模式，它将中心化的数据管理与分布式的、轻客户端的数据审计相结合，从而兼有中心化的高效便捷性与去中心化的安全透明性。

3 类区块链存储

类区块链存储的目的是利用区块链的数据结构“锁住”数据，一方面防止中心随意篡改和伪造数据，另一方面也为下文的分布式审查提供可查验的凭证。在形式上，传统 CA 模式直接存储业务数据本身，只有数据没有链；而在基于区块链技术的 DA 模式中，所有的业务数据都存储在链上，数据就是链，链就是数据。本文的 SDA 模式将数据与链分离，业务数据像 CA 模式一样由中心独立存储，同时在公开的链上存储与业务数据相关的验证数据。

具体而言，SDA 模式中数据的存储分为两部分，中心数据和客户端数据。其中，中心存储的内容包括：（1）由业务数据构成的中心数据库 CDB；（2）记载每一条记录的哈希摘要的区块链；（3）每一个区块的 Merkle 可信树。客户端存储的内容包括：（1）由记录哈希摘要的零星副本片段构成的本地数据库 LDB；（2）用户核验记录存在性的总账本——块头哈希链。

3.1记录设计

SDA 模式中的数据记录有两类，一类是中心存储的完整数据R_{CDB}，另一类是客户端存储的摘要数据R_{LDB}。

中心记录的用途是满足用户查询和验证的需求，因此除完整的业务数据外，还须有少量用于验证和检索的附属数据。

$$R_{CDB}: \mathrm{ID}, \mathrm{H(Key)}, \mathrm{Data}, \mathrm{Es}\big(\mathrm{ID}, \mathrm{H(Key)}, \mathrm{H(Data)}\big)$$

其中，Data 表示应用的业务数据；Key 表示每一条记录的关键字，H(Key)为其哈希值，它是用户查询检索和集合审查的依据，同时进一步保护数据隐私；ID 表示记录号，它由区块号和块内顺序号两部分组成，是用户更新下载和集合审查的依据，如 B101#2 表示 101 号区块的第 2 条记录；H(Data)表示数据摘要，它是业务数据的哈希值；$\mathrm{Es}(\mathrm{ID}, \mathrm{H(Key)}, \mathrm{H(Data)})$表示中心对ID & H(Key) & H(Data)三部分数据的联合签名，它是经过中心签名的完整数据指纹，简称Es(R)。需要说明的是，ID 和H(Key)都是记录的候选码，具有全局唯一性，分别用于不同的查验方式。

客户端记录的用途是查验中心数据的真实性，因此R_{LDB}无需保留业务数据本身。

$$R_{LDB}: \mathrm{ID}, \mathrm{H(Key)}, \mathrm{Es}(\mathrm{ID}, \mathrm{H(Key)}, \mathrm{H(Data)})$$

这样的设计兼顾了数据的隐私性与可监督性，在保证中心数据隐私的前提下，既提供了集合查验的数据基础，也极大的降低了客户端的存储负担。

3.2 区块链设计

如前所述，SDA 模式引入块链架构的目的是锁住数据，供用户验证中心每一条记录的存在合法性。因此，区块体中存储的并非记录 R 本身，而仅仅是其签名指纹Es(R)。新区块的生成既可以采用数量模式，如每个区块固定加载 1024 条记录；也可以采用时间模式，如周期性的定时生成区块。本文以中心每天定时生成一个区块为例，当前区块包含前一天所有新增记录的Es(R)值，且对应唯一的“时间戳”。同时，用 Merkle 可信树来归纳区块中的所有Es(R)值，生成整个区块的数字指纹，从而提供了一种校验记录 R 是否存在于某个区块的高效途径[10]，如图 2 所示。[1]

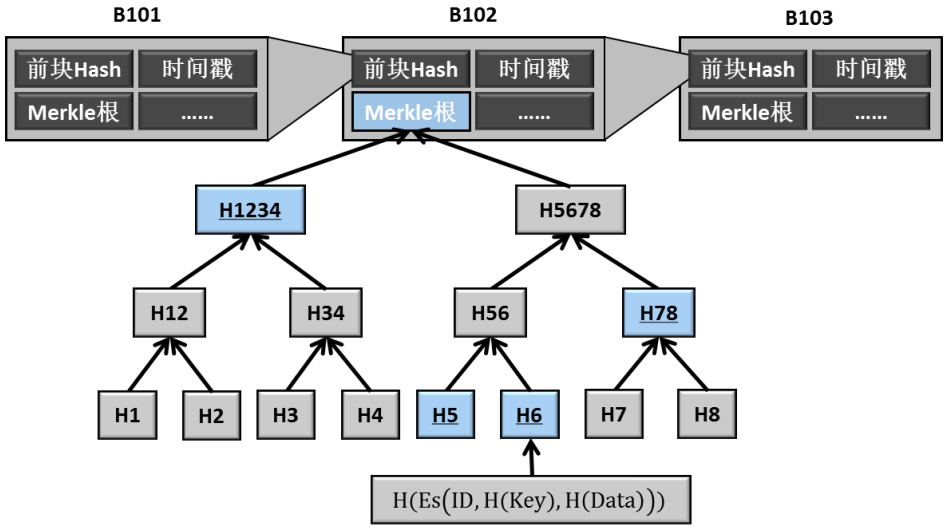

Fig. 2　The example of a Merkle tree included in a block, which contains 8 records.

图 2　基于 Merkle 可信树的存在性证明（一个 8 记录区块的示意图）

所有区块按照时间先后顺序构成一条证据链，如图 3 所示。每一个区块的块头记载的内容除本块的核心信息外（如 Merkle 根、记录总数、时间戳等），还包含一个特殊字段——经中心签名的“父区块的哈希值”Es(H(Pre_Block_Head))（准确而言，签名对象是前一个区块的区块头的哈希值），从而锁定该区块的唯一前导区块。[10]同时，中心签名保证了数据来源的可靠性，并为事后质证提供了依据。显然，区块链具有“牵一发而动全身”的特点，历史记录按区块顺序被依次唯一锁定，保证了所有记录的完整性、可追溯性和不可篡改性。

---

[1] Merkle 树是一种哈希二叉树，它是一种用作快速归纳和校验大规模数据完整性的数据结构。构造一棵完整的 Merkle 树需要递归的对数据节点进行哈希运算，并将新生成的哈希节点插入到 Merkle 树中，直到只剩下一个哈希节点，该节点就是 Merkle 根。
当 N 个数据元素经过哈希计算并插入 Merkle 树时，任意一个满节点回溯至多 $\log_2(N)$个路径节点就能到达 Merkle 根，即至多经过 $\log_2(N)+1$ 次哈希计算就能检查出任意数据元素是否存在于该 Merkle 树，从而完成其存在性证明。

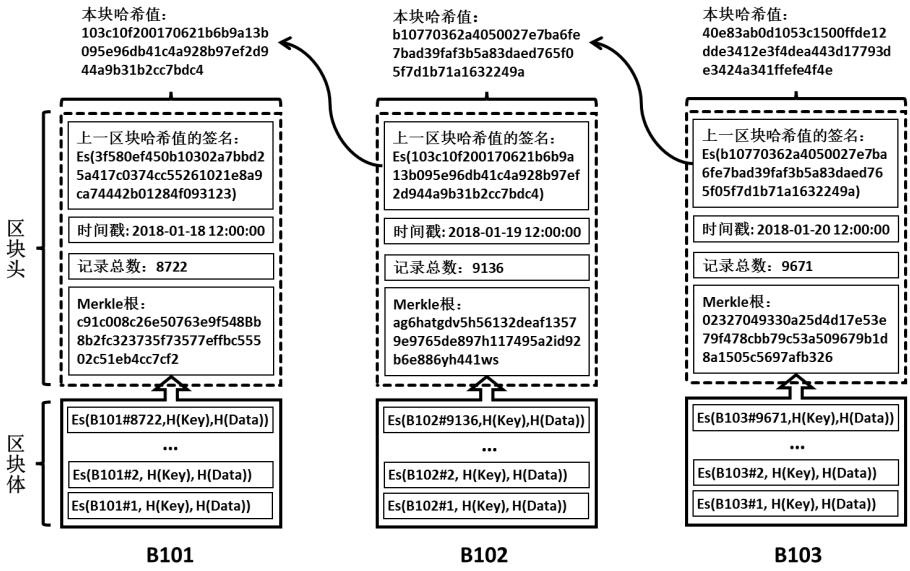

Fig. 3　An example of the blockchain used in the proposed architecture.

图 3　类区块链的示意图

4 分布式集合查验

分布式审查的目的是防止中心自身作假，以保证数据的真实性。数据的真实性体现在三个方面：来源可靠，真实存在，独一无二。其中，利用类区块链存储技术，我们可以顺利查验前两个方面的真实性，即中心的哈希签名保证了数据来源的真实性，Merkle 可信树保证了数据存在的真实性。因此，现在的关键问题是如何在单一用户不掌握完整数据，只有一些零星的记录指纹副本的情况下，通过群体协作来验证每一个关键字对应的记录在中心只存储有一个版本，即中心没有伪造的同名记录（相同 Key 值）。SDA 模式通过大量客户端随机集合抽查的方式，巧妙的解决了这一问题。

所谓“分布式集合查验”是指大量的、分散的客户端利用自身已下载的本地 LDB 数据和块头哈希链，以记录集合方式整体性查验，在数据查询的同时完成对中心数据的验证和本地数据的更新。通过大量个体的分散查验，达到整体普查的效果，从而完成对中心的民主监督。

每个客户端保存有完整的“块头哈希链”——由所有区块的区块头构成的数据链，它们是每个区块内所有记录的整体性数据指纹。每次查询前，客户端先更新块头哈希链，即下载自上次更新以来中心新增的区块头数据，并依据先前保留的数据逐步验证新增块头数据的合法性与延续性。对于用户的查验请求，中心必须响应。若因网络故障，客户端会固执的反复发送相同的数据请求，直至得到回复。此外，为确保查验算法的公开透明，客户端程序必须公开源码。

在具体实施中，分布式集合查验主要用到三项关键技术：双盲抽查、集合验证、查询/验证/更新相结合。

（1）双盲抽查，它是指中心在应对客户端的查验时既不知道用户的具体身份，也不知道集合查验中哪些记录是用户已知的，哪些是用户未知的。双盲抽查避免了中心选择性作弊，即中心根据客户的查验历史推导客户端 LDB 的记录内容，从而有针对性的对客户未知记录作假。

（2）集合验证，它是指客户端对中心采取一次性批量查验的方式，而非一条一条的多次查验。集合验证大大提升了记录的查验数量和速度，并大幅降低了中心与客户端之间的通信量，从而提升了系统的查验效果，降低了中心作弊被漏查的概率。

（3）查询/验证/更新三结合，它是指查验机制将记录查询、记录验证和本地 LDB 记录更新结合起来，三位一体，三步工作一次性完成，进一步提升了查验效率，同时避免了中心单独在查询环节或更新环节作弊。

下面详细阐述 SDA 模式的分布式集合查验算法，以及本地 LDB 的更新策略。

令待验证记录集合[R] = $[R^*_{LDB}] \cup [R_{new}] \cup [R_{query}]$。其中，$[R^*_{LDB}]$表示客户端 LDB 中已知记录的随机子集，它是验证集合的主体，占比通常在 90%以上；$[R_{new}]$表示 LDB 之外未知记录的随机子集，它是待下载更新部分，只知道 ID 不知道 Key 值(或H(Key))；$[R_{query}]$为待查询记录集，只知道 Key 值不知道 ID。完整的查验算法如下（基于 C/S 模式）：

（1） C：更新块头哈希链，并结合自身已有的块头数据逐步验证新增区块的合法性；

（2） C->S: $50\%\,\mathrm{H}(R^{key}),\ 50\%\,R^{ID}$

客户端匿名[2]发送待查验记录集合的检索字串，两种检索方式各占一半，检索值分别按照H(Key)和 ID 降序排列，其中$[R_{new}]$以 ID 方式检索，$[R_{query}]$以H(Key)方式检索，$[R^*_{LDB}]$按剩余比例随机表示为H(Key)或 ID。

（3） S->C: $V_{Es} = \mathrm{Es}(\ \mathrm{H}(\cup \mathrm{H}(R^{key})),\ \mathrm{H}(\cup \mathrm{Es}(R)))$

中心的回传值是对两个哈希值的联合签名，这两个哈希值分别是待查验记录的关键字哈希值的联合哈希和签名指纹的联合哈希[3]。需注意的是，联合哈希值的顺序必须与步骤（2）中[R]的检索字串顺序相一致。[4]

（4） C->S: $R^{ID}_{new},\ R^{key}_{query}$

客户端明确指定待查询记录的 Key 值和待更新记录的 ID。

（5） S->C: $R^{ID}_{new}, H(R^{key}_{new}), \mathrm{Es}(R_{new}), \mathrm{Merkle}(R_{new});\ R^{Data}_{query}, \mathrm{Merkle}(R_{query})$

中心回传待更新记录集$[R_{new}]$的 LDB 数据和 Merkle 路径，以及待查询记录集$[R_{query}]$的完整数据和 Merkle 路径。[5]

（6） C: $\mathrm{Verify}(R_{new}), \mathrm{Verify}(R_{query}), \mathrm{Verify}(V_{Es})$

客户端利用步骤（3）、（5）的回传数据和本机 LDB 数据及块头哈希链依次验证$[R_{new}]$和$[R_{query}]$各记录的身份合法性和存在合法性，以及V_{Es}值的正确性。

（7） C: $Update(R_{LDB})$

客户端更新本地数据库，将$[R_{new}]$和$[R_{query}]$的相关数据添加到 LDB。

客户端本地 LCD 的更新策略：$[R_{new}]$中未知记录的选择，客户端根据块头哈希链的信息随机生成区块号和块内顺序号，构成待更新记录的 ID。其中，随机算法对区块的选择略微具有时间偏向，尤其对最近的几个新区块具有较高的选择倾向，以确保新区块中所有记录的签名指纹能够被充分的分散保存，进一步防范中心在新区块中伪造同名记录副本。

5 效率分析

SDA 模式结合“类区块链存储”技术和“分布式集合查验”机制来维系数据共识，通过三个合法性检验保证了数据的真实性：（1）数字签名提供了记录的身份性证明，即数据来源可信任，并提供了质证的依据；（2）Merkle 可信树提供了记录的存在性证明，确保中心无法篡改原始记录；（3）最后，分布式集合审查提供了记录的唯一性保障，确保中心无法将原始记录调换为伪造的同名记录副本。

在上一节的查验算法中，步骤（2）和（3）是关键。它们要求中心在不知道用户真实查询意图的情况下，计算包括查询、验证、更新三类记录集合的联合数据摘要值。如果中心企图作假，只要有一条虚假记录落在$[R^*_{LDB}]$中，中心就无法通过步骤（6）的Verify(V_{Es})测试。由于$[R^*_{LDB}]$是每次查验的绝对主体，基于双盲机制的集合查验确保了单次漏查的概率极小。

---

[2] 为防止中心因匿名查询遭到 DDoS 攻击，可以在每次查询前进行一次 CAPTCHA 验证码测试，从而有效避免大规模机器的恶意查询。

[3] 记录集合的关键字哈希值的“联合哈希”是指将集合中每条记录的关键字哈希值H(Key)依次连接，构成一个联合字串，然后对该联合字串求哈希。“签名指纹的联合哈希”类似。

[4] 若$[R_{query}]$中待查询的个别记录不存在，中心 S 应返回缺失记录的 Key 值以及除此之外其他记录的V_{Es}值。

[5] 如有必要，可以在步骤（3）后增加用户身份合法性检验，并在步骤（5）中加入存取控制检验以强化数据安全，核查客户身份及合法权限，即该用户是否有权限查询指定记录。

$$P(\textit{中心作弊被漏查}) = \frac{crad([R_{new}] \cup [R_{query}])}{crad([R^*_{LDB}] \cup [R_{new}] \cup [R_{query}])}$$

即使中心作假侥幸躲过某次查验，由于更新记录$[R_{new}]$和查询记录$[R_{query}]$都会存放到用户 LDB 中用于以后的查验，所以中心只要对某条记录有过一次作假，就必须在以后的查验中对它一直作假，否则会以大概率露馅。然而，总会有保留真实记录Es(R)值的用户[6]，他们的查验会百分百的暴露中心作假的事实。因此，整体而言，中心作假被漏查的概率为零。

同时，客户端获取的所有查验数据和块头哈希链都经由中心签名，具有不可抵赖性。只要坐实一次造假，就会危及对中心的信任，直至摧毁整个应用体系，即中心作假的成本无限高。所以，SDA 模式的三个合法性检验机制保证了中心只能真实公正的提供数据服务，不能、不敢也不愿虚假操作。

SDA 模式采用的是一种轻客户端验证方式。我们以一个 10 亿条记录（230）的超大型数据中心为例，假设每个区块装载 216 条Es(R)，则共有 214 个区块，Merkle 树的存在性证明至多需要 16 个哈希路径节点，共 17 次哈希运算。若每天生成一个区块，需要约 45 年的时间。若采用 DA 模式，这样的数据规模即使每条记录只有 1KB，矿机仅数据存储空间就超过 1TB。而在 SDA 模式下，客户端 LDB 中签名指纹的大小与记录本身的大小无关，每条R_{LDB}不超过 100B；每个区块头的大小同样与区块体大小无关，每个 BlockHead 不超过 100B。若每个用户的 LDB 中有 10 万条Es(R)，每次查验集合为 1000 条，其中 990 条取自本身 LDB，则用户单次漏查概率为 1%。用户块头哈希链的大小不超过 16MB，LDB 的大小不超过 10MB，单次通信量除待查询数据R^{Data}_{query}自身外，额外数据不超过 32KB[7]，哈希计算约 172 次。显然，对于普通电脑或手机客户端而言，无论存储量、计算量还是网络带宽都能够轻松胜任。

以上分析表明，SDA 模式对于客户端而言是一种轻量级的、高效的、可信的数据应用体系。而对于中心而言，对比 CA 模式主要的成本是单用户查询次数增大了约 1000 倍，除此外其他的计算和网络负担都可以忽略不计，而这种本地化的简单查询对中心而言也完全是可接受的。

6 结论

区块链应用的核心价值并不是去中心化，而是构建了可被监督的价值流通体系。在区块链领域，如何利用其技术特性构建一个相对平等的机制，让用户和平台能够相互监督，才是比较贴近实际的想法。去中心化只是其中一种方式，本文提出了另一种更加高效的方式。

本文设计的混搭架构的数据应用体系结合中心化的数据存储模式与去中心化的数据共识机制，实现了一种信用机制的创新。它在数据应用模式上首创了一种区别于区块链完全分布式管理和传统集中式管理之外的第三条路径——“半去中心化”模式，即一方面保证数据公开与透明，具有区块链的不可篡改性和可追溯性特征，同时数据集中存储，满足了政府或大企业自己掌控数据，不完全公开数据的需求。在这种模式中，数据的公信力既不依赖于强力第三方的信用背书，也不依赖于无数完全冗余副本的低效共识，而是根源于大量客户端轻负载的分布式民主监督，并且很好的兼顾了隐私与监督之间的平衡。

半去中心化的数据应用体系并非针对某个具体应用而设计,具有广泛的应用前景，其上可存储日志、文件、合同、邮件、财务凭证、聊天记录等任何数字化信息，用于打造一款应用广泛的、可信任的“数据银行”，成为未来数字普惠金融的重要基础设施。

Reference:

[1] Bakir V. Policy Agenda Setting and Risk Communication Greenpeace, Shell, and Issues of Trust[J]. The Harvard International Journal of Press /Politics, 2006, 11(3): 67-88.

[2] Andrew M., Ari J., Elaine S., et al. Permacoin: Repurposing Bitcoin Work for Data Preservation[C]. Proceedings of IEEE Symposium on Security and Privacy. Washington D C IEEE, 2014: 475-490.

[6] 按照查验算法中 LDB 的更新策略，新区块记录的签名指纹会被优先下载，因此几乎每一条记录的 Es(R) 值都会分散到若干用户的 LDB 中。即使某条记录以极小概率一直未被下载，未来也随时可能在某次查验中跟随$[R_{new}]$集的随机 ID 号被更新到某个用户的 LDB 中。

[7] 通信量主要集中在步骤（2）的联合查验字串，其次是步骤（5）每条未知记录的 Merkle 树路径节点的哈希值。

[3] Binanda S., Samiran B., Sushmita R., et al. Retricoin: Bitcoin based on compact proofs of retrievability[C]. Proceedings of the 17th International Conference on Distributed Computing and Networking. New York: ACM, 2016, 14: 1-10.

[4] Juels A., Kaliski B.S. PORs: Proofs of retrievability for large files[C]. Proceedings of ACM Conference on Computer and Communications Security. New York: ACM, 2007:584-597.

[5] Walsh K. 'Blockchain' Increases Online Trust: New Technology Could Bolster Digital Badges and E-Portfolios, among Other Innovations[J]. University Business, 2017, 20(2): 24.

[6] David S. Understanding the DAO Attack[EB/OL]. https://www.coindesk.com/understanding-dao-hack-journalists/, 2016.6.

[7] Pete R. Split or No Split? Bitcoin Miners See No Certainty in Segwit2x Fork[EB/OL]. https://www.coindesk.com/split-no-split-bitcoin-miners-see-no-certainty-segwit2x-fork/, 2017.11.

[8] Xiwei X., Cesare P., Liming Z., et al. The Blockchain as a Software Connector[J]. Software Architecture, 2016: 182-191.

[9] Roman B., Michel A., Matti R., et al. Blockchain Technology in Business and Information Systems Research[J]. Business & Information Systems Engineering, 2017, 59(6): 381-384.

[10] Andreas M.A. Mastering Bitcoin[M]. O'Reilly Media, 2014.12.