专刊:数据要素安全高效流通的关键技术
Special Issue: Key Technologies for Safe and Efficient Circulation of Data Elements

ISSN 2096-742X
CN 10-1649/TP

文献CSTR:
32002.14.jfdc.
CN10-1649/
TP.2023.05.005

文献DOI:
10.11871/jfdc.issn.
2096-742X.2023.
05.005

页码:46-62

获取全文

基于区块链的数据要素可信流通技术综述

钟子岳, 朱长昊, 李浚哲, 张美慧*

北京理工大学,计算机学院,北京 100081

摘　要: 【背景】数据已然成为经济发展的基础性战略资源。要充分发挥数据要素作用,需要建立数据可信流通体系。【目的】区块链作为新型可信数据管理平台,能够实现数据可信流通的基本需求。因此,本文将探讨基于区块链的数据要素可信流通技术。【方法】从建立数据可信流通体系的角度出发,本文首先分析了构建数据要素可信流通体系的基本要求,然后分析了基于区块链实现数据可信流通的技术要点,总结了目前区块链研究领域中可用于实现上述目标的研究工作,并对未来的研究方向提出展望。【结果】着眼于增强数据可用、可信、可流通、可追溯4个方面,本文对现有区块链系统研究中的存储模型、系统性能扩展方式、数据验证、跨链技术、溯源技术等方面进行总结分析。【结论】现有研究成果能够基本实现数据要素可信流通体系,但在数据隐私、数据使用和数据表示方面仍存在诸多未解决的问题。

关键词: 区块链;数据要素流通;数据管理;数据存储;分片;跨链;溯源

A Survey on Blockchain-Based Trusted Data Elements Circulation

ZHONG Ziyue, ZHU Changhao, LI Junzhe, ZHANG Meihui*

School of Computer Science & Technology, Beijing Institute of Technology, Beijing 100081, *China*

Abstract: **[Background]** Data has become a kind of fundamental strategic resource for economic development. To fully leverage the value of data elements, a trusted data circulation system needs to be established. **[Objective]** As a new type of trusted data management platform, blockchain can realize the basic needs of trusted data circulation. Therefore, we discuss the trusted circulation technology of data elements based on blockchain in this survey. **[Methods]** From the perspective of establishing a trusted data circulation system, we first analyze the basic requirements of building a trusted circulation system for data elements. Then, we analyze the main techniques for realizing trusted circulation of data based on blockchain. Finally, we summarize current research works in the research field of blockchain that can be utilized to achieve the mentioned goals and present prospects for future research directions. **[Results]** Focusing on the aspects of enhancing data availability, credibility, negotiability, and traceability, we summarize and analyze the storage model, system performance scaling methods, data verification techniques,

基金项目: 国家自然科学基金(62072033)

*通信作者: 张美慧(E-mail: meihui_zhang@bit.edu.cn)

cross-chain techniques, and data provenance techniques of the existing blockchain system. **[Conclusions]** The existing methods can realize the trusted circulation system of data elements to some extent, yet there are still several unresolved problems in data privacy, utilization, and description.

Keywords: blockchains; circulation of data elements; data management; data storage; sharding; cross-chain; provenance

引　言

数据具有推动经济发展和产业变革的重大作用。从2017年到2021年,我国数据年产量从2.3ZB增长至6.6ZB,全球占比9.9%,位居世界第二[1],然而,接近70%的数据价值未被激活。2019年10月,党的十九届四中全会将数据纳入生产要素范畴,其原因在于它对推动生产力发展显现出的价值[2],是对数据价值定位不断深化的体现,强调了数据的重要性。然而,数据显著推动生产需要相应的技术和产业基础[2],并且与传统生产要素不同,数据要素流通涉及个人隐私及安全[3],且具有可复制、可共享、无限增长和供给的特点[2],对数据管理提出新的需求。国务院《关于构建数据基础制度更好发挥数据要素作用的意见》中指出:构建数据基础制度,要建立数据可信流通体系,增强数据的可用、可信、可流通、可追溯水平[4]。

区块链是一种去中心化的新型数据管理平台[5],在不可信网络中进行可信的数据存储与事务处理。传统区块链系统可以在一定程度上满足数据要素管理中的可用、可信、可流通、可追溯的要求:

(1)在数据可用性方面:区块链系统是一种由多个节点运行的冗余系统,数据被分布式地存储在众多节点上。在传统的区块链系统中(例如:比特币[6]、以太坊[7]、超级账本[8]等),系统中的每个节点都存有完整的数据副本,即使某些节点出现系统故障或网络故障,也不影响整个区块链网络的运行,保证了数据的冗余备份和可靠性。

(2)在数据可信性方面:区块链系统是一种去中心化的分布式系统,无中心控制机构,避免依赖于机构或个人的信用背书。通过共识机制,区块链网络中的每个节点都有权查看和验证区块链上的数据,数据一旦经过验证并被添加到区块中,将无法被单一或小部分恶意节点篡改和删除。这种去中心化、不可篡改、不可删除的特性保证了系统中数据的高度可信。

(3)在数据可流通方面:区块链系统可以建立在多个互不信任的参与方之间,使数据能够在不同组织和系统之间自由流通,打通了传统系统之间的信息壁垒。

(4)在数据可追溯方面:区块链将所有的交易记录按照时间顺序写入到去中心化账本中,任何参与方都可以根据账本记录追溯数据的流转途径。每笔交易都使用数字签名进行认证,从而确保交易的参与者是真实的,交易的内容是不可篡改、不可否认的。

与分布式数据库相比,区块链在安全和可流通性方面具有优势,但是区块链系统在数据管理方面仍面临诸多问题,阻碍了区块链系统成为建立数据可信流通体系的核心。幸运的是,区块链和分布式数据库系统存在很多相似的技术概念和解决方案,使双方在安全、效率和隐私方面的优势得以结合[9]。本文将从区块链系统的角度,综述用于建立数据要素可信流通体系的关键技术。本文将从数据的可用、可信、可流通、可追溯4个方面综述现有区块链系统在数据管理方面的研究工作:

(1)在数据可用性方面:传统区块链系统主要采用键-值型存储模型。数据要素场景中包含金融监管、数字存证、政务服务等多种应用场景,需要更加具有表达能力的存储和查询引擎。此外,相较于传统分布式数据库,区块链系统存储的可扩展性较差,难以面向大数据应用场景。因此,本文将针对现有区块链与数据的融合研究中用于提升区块链数据可用性的关键技术进行讨论。

(2)在数据可信性方面:现有区块链系统在数据查询验证方面主要采用两种策略。一种是

基于密码学技术的方式,通过全网共识的可验证数据结构(Authenticated Data Structures, ADS)提供查询证明,另一种是基于可信硬件保证事务执行可信。本文主要针对现有系统中ADS所支持的查询种类和验证性能,以及基于可信硬件的事务执行方式进行讨论。

(3)在数据可流通方面:在实际应用中,不同的业务系统往往采用相对独立的区块链系统,因此便产生了多链环境。然而,链与链之间的共识协议并不相通,数据账本也不共享,这就产生了数据孤岛问题。区块链跨链技术可以实现不同区块链之间的安全、高效的数据传递,进一步促进了数据的交换和流通。区块链跨链技术仍然是区块链研究领域中的一个热点话题,因此本文将对目前研究领域中多种区块链跨链技术的机制和难点进行讨论。

(4)在数据可追溯方面:区块链的去中心化及保留全部历史数据的特点为数据可信追溯提供支持。然而,传统区块链系统上数据回溯的效率低,溯源结果及演变过程的真实性难以验证,溯源数据上链过程缺乏有效管理。本文将对目前区块链研究工作中用于提升溯源查询和验证效率及保证溯源信息的准确性两个方面的研究进行总结与讨论。

1　区块链技术背景

区块链的概念起源于中本聪提出的比特币[6],其本质上是一种把数据块按照时间顺序连接而成一种链式数据结构。以比特币为例,区块在网络中所有节点以一致的方式顺序连接并存储,新的区块只能被添加到链表的末尾。在此过程中,区块是区块链的基本结构单元,由包含元数据的区块头和包含事务数据的区块体组成。具体来说,区块头由区块高度、前一个区块的哈希值、时间戳、nonce、矿工签名和Merkle Tree根哈希等信息组成,而区块体可以被视为由多个事务组成的事务记录的集合。在比特币中,区块由哈希值连接,区块中数据的任何变化都会导致该区块的哈希值发生变化,进而影响到该区块之后所有区块的有效性,因此在区块链中篡改数据会随着区块链的增长而愈加困难,保证了数据存储的安全性。

随着人们对区块链的理解不断加深,目前的区块链创新融合了密码学、P2P通信、共识机制、智能合约等多种计算机前沿技术,已经发展成为利用链式数据结构来验证与存储数据、利用分布式共识算法来生成和更新数据、利用密码学的方式保证数据传输和访问的安全、利用由自动化脚本代码组成的智能合约来编程和操作数据的一种全新的分布式基础架构与计算范式[10],具有分布存储、不可篡改、可追溯、可编程等技术特征,构建了独特的信任机制,在数据可信流通领域得到了广泛应用。

而在功能架构方面,区块链系统可以自下而上分为以下5层,为区块链的各种功能奠定基础:

(1)数据层。为了有效地组织区块链中的各种数据,数据层包含数据结构、事务模型、索引数据、状态数据和持久存储方案等元素。

(2)网络层。为了满足去中心化网络中各节点之间的通信,P2P协议在网络层起着重要的作用。节点间传输的内容主要由事务数据和区块数据组成。

(3)共识层。与中心化管理的数据库不同,区块链使用分布式共识算法来确保网络中互不信任的节点可以在同一账本上达成一致。共识算法的使用提高了区块链应对崩溃容错或拜占庭容错的能力,使区块链比传统数据库具有更高的安全性。

(4)合约层。包含各种脚本、算法和智能合约,是区块链可编程性的基础。

(5)应用层。用户可以使用区块链提供的API很容易地开发新的去中心化应用程序。

2　区块链数据可用性

区块链技术承担着数据要素可信流通过程中的存储、交易与共享等关键环节,主要支撑数据存储和查询功能。本文从区块链的存储模型

和可扩展性两个方面来探讨区块链的数据可用性。存储模型是系统设计中至关重要的层次,与查询密切相关;可扩展性是系统能否面向大数据应用场景的重要技术指标。

2.1 区块链的存储模型

早期的区块链系统主要应用于加密数字货币[6-7]及金融服务领域[11],用于存储数字资产。这些系统主要使用键-值型存储模型,其优势在于:(1)键-值型存储模型足以满足以账户模型为主的数字资产存储需求;(2)键-值型存储模型通常不需要复杂的数据结构来存储数据,能够带来较高的查询处理性能。

随着区块链在数据管理领域逐渐得到人们的关注,广泛应用于金融监管、数字存证、政务服务和溯源防伪等场景中,单一的键-值型存储模型已经不能满足多样化的业务需求,存在查询语义缺失、查询能力不足等问题[12]。研究者们开始探索在区块链系统中实现数据库的存储模型。目前的研究主要分为两种方式:一种方式是不改变区块链原有的存储引擎,通过改变数据的存储格式,并引入多种索引,在区块链系统之上构建了支持各种查询功能的抽象数据库层;另一种方式是将成熟的数据库系统作为区块链的新型存储引擎,并修改区块链所提供的查询处理接口,从而使区块链系统具备数据库的查询处理能力。

SEBDB[12,13] 属于第一种实现方式,将关系型语义添加到区块链存储中,通过重新定义区块链交易的数据格式实现了一种表格存储模式,并将每条交易作为一项数据记录。具体来说,交易包含两种属性类型:系统级属性和应用级属性。系统级属性包括交易ID、签名、时间戳、交易发送者、交易类型等,由系统自动添加,其中交易类型用于区分该条记录所属的表格。应用级属性由用户在表格创建时进行显示定义。为提升链上数据的查询效率,SEBDB定义了3种基本查询操作:(1)根据区块号、交易记录编号或时间戳来获取区块数据;(2)根据交易类型获取全部交易记录;(3)获取满足特定条件的交易记录。随后提出3种索引来优化上述查询:(1)以(block_id, first_tx_id, ts)为键的区块级B+树,用于根据给定区块号、交易ID或时间戳来获取区块位置。(2)表级位图索引,索引中的每一行代表一个表格,每一列代表一个区块,第i行第j列表示区块j中是否包含表格i中的交易记录。(3)分层索引,其中第一层使用位图索引来描述属性值在区块之间的分布,第二层使用B^+树索引块内属性数据。通过上述3种基本查询操作和索引,SEBDB实现了简单的CREATE、INSERT、SELECT、JOIN等部分SQL查询操作。

BRD[14] 和 FalconDB[15] 属于第二种实现方式。BRD直接使用PostgreSQL[16]作为区块链的存储引擎,实现了第一个具有区块链属性的去中心化关系型数据库。基于可串行化快照隔离机制(Serializable Snapshot Isolation, SSI),BRD实现了事务的并发执行,并通过排序服务将事务在不同节点上以相同的可串行化顺序提交,保证区块链节点间的一致性。FalconDB使用IntegriDB[17]作为区块链的存储引擎。其中,IntegriDB是一种可验证的外包数据库,利用MySQL[18]进行数据存储,并提出基于密码学累加器的ADS存储并生成数据查询验证证明。FalconDB在数据表中记录完整历史数据,为每个数据表自动添加VF(valid from)和VT(valid to)两个属性,用于表示数据的可见范围。数据被创建或修改时,将操作事务所属的区块号写入VF字段,数据被删除时,将删除事务所属的区块号写入VT字段。因此,对于区块高度为h的数据版本,只有$VF \leqslant h$且$VT \geqslant h$的数据记录是可见的。

2.2 区块链存储的可扩展性

由于区块链数据以全副本的形式保存在所有节点中,导致链上数据存储成本高,共识开销大,影响了区块链的可扩展性。为提高区块链系统的可扩展性,现有的研究工作主要分为链下方案和链上方案两种实现方式[19]。前者的核心思想是将数据管理系统中的大量存储和计算转移至链下,链上只保留必要的验证数据。后者主要

以区块链分片方案为主,旨在提高传统区块链的可扩展性,这种方式参考了分布式数据库的思想,将区块链网络中的节点划分为若干子集,并将存储和计算划分到不同子集中,实现存储和计算负载均衡的同时,降低网络压力。

2.2.1 链下扩展方案

在链下扩展方案的研究工作中,SEBDB率先实现了链上链下协同处理,基于为区块链扩展的关系型语义,提出了链上-链下连接(On-off Join)操作,对区块链上的表格和链下数据库中的表格进行连接查询,从而支持方便的应用程序开发和有效的复杂任务建模能力。

徐建良等人在多篇工作中提出链上链下混合存储架构[20-21],致力于解决数据在区块链系统中存储开销大(例如:以太坊的Gas花费)的问题。其主要思想是使用更多的链上读操作和链下计算操作代替链上更新操作,减少ADS更新过程中导致的链上存储开销。其中,GEM^2-tree主要以若干棵Merkle B-tree(MB-tree)[22]组成,在链下存储中保存完整、有序的数据和可验证数据结构MB-tree,在链上状态数据中以仅追加的方式无序存储数据,并以最小存储成本保存MB-tree根哈希用于数据验证,目的是减少链上数据排序过程中的存储开销。Chameleon^{inv}索引[21]利用变色龙向量承诺技术(Chameleon Vector Commitment, CVC)使ADS的更新不需要修改变化节点的父节点保存的数据摘要,在ADS更新过程中无须修改树根摘要,从而比GEM^2-tree进一步减少了链上数据的更新操作,极大地降低了混合存储系统的开销。

LightChain[23]则结合区块链和分布式哈希表形成链上链下存储方案,从两方面提升了区块链的可扩展性。一方面,该架构将原始数据存储在分布式哈希表DHT中,使用DHT生成的哈希值在区块链上进行交易;另一方面,区块链中的每个节点无须存储所有的区块数据,只需存储随机分配的一部分区块。

2.2.2 链上扩展方案

区块链分片方案主要包含三部分内容:网络分片、计算分片和存储分片。其中,网络分片是基础,存储分片是瓶颈[24]。网络分片是通过将节点划分为不同的共识子集,从而降低共识过程中的网络压力。计算分片是将交易的处理划分到不同的分片,从而降低节点的计算压力。存储分片是将数据存储(包括区块数据或状态数据的存储)划分到不同的分片,从而降低节点的存储负担。

Elastico[25]是首个分片区块链系统,将网络中的节点划分为不同子集,子集内分别独立运行PBFT协议进行共识,没有考虑子集间的跨分片事务处理。OmniLedger[26]引入了2PC协议用于跨分片事务处理。Monoxide[27]提出一种诸葛连弩挖矿(Chu-ko-nu Mining)解决了PoW共识协议在网络分片后导致的算力分散问题,以增强系统安全性。BrokerChain[28]通过将账户拆分为多个子账户,并划分到不同分片中,解决了资产交换场景下的跨分片事务处理问题。由于无许可区块链(Permissionless Blockchain)的节点信任度较低,研究工作在提升区块链可扩展性的同时,还要着重保障分片数据的安全性。为防止节点作恶行为发生,通常要设置分片纪元。每个分片纪元结束后对节点进行分片重新配置(Reconfiguration),分为二次全随机分配[25]、部分重分配[29]和自由选择重分配[30]等方式。

而许可区块链(Permissioned Blockchain)通常不需要重新配置环节,研究工作主要集中于提升区块链系统的事务处理能力。目前,大多数区块链分片的研究工作能够很好地解决网络分片和计算分片的问题。例如:Amiri等人[31]提出了一种用于许可区块链分片的账本模型,该模型是一种有向无环图DAG形式的账本,每个节点都存储与其相关的账本视图,跨分片事务由多个相关分片共同存储。在这种架构下,非跨分片事务和没有读写冲突的跨分片事务之间可以并发执行[32]。然而,在跨分片事务比例较高时,该系统不能表现出良好的事务处理性能。

因此,存储分片仍然是亟待解决的问题,主要体现在跨分片事务处理的性能问题,尤其是通用数据查询事务的跨分片处理。Pyramid[33]使用桥梁分片来处理跨分片事务。具体来说,Pyra-

mid中的节点分为i-shard和b-shard两种。i-shard存储单个分片数据,可以单独执行片内事务。b-shard存储多个分片的数据,作为跨分片事务的桥梁,可以直接执行跨分片事务执行,并协调相关i-shard进行分片数据更新,从而提升跨分片事务的处理效率。Meepo[34]将分片下放到区块链存储层,区块链系统中的每个参与方都保存完整数据账本,每个参与方包含多个节点用于进行数据分片。在这种方式下,跨分片事务将被限制在每个参与方内部,避免了信任问题,从而可以使用故障容错协议(Crash Fault Tolerance, CFT)。然而,由于每个参与方都需要进行多节点部署,该系统对参与方的硬件要求较高。GriDB[35]是首个支持SQL查询的可扩展区块链数据库,将跨分片事务通过关系代数分解后。借鉴链下支付的思想,GriDB将数据操作分别委托给各个分片中的一个节点进行,再将结果汇总到主分片,并通过多种新型ADS实现跨分片数据验证。GriDB还以同样的思想设计了跨分片数据迁移机制,用于动态调整数据表格所存放的节点,以实现负载均衡。

2.3　讨论

本节介绍了区块链存储模型和可扩展性两方面对区块链数据可用性的影响。表1对这些系统进行了总结对比。

在存储模型方面,键-值型存储模型适用于工作负载简单,数据检索性能要求高的场景,而关系型存储模型适用于复杂工作负载,更贴近数据要素共享流通场景。

在可扩展性方面,将庞大的数据存储和计算迁移到链下进行能够显著提升区块链系统的处理能力,而传统的链上扩容方案难以从根本上解决系统性能问题,主要原因在于跨分片事务的处理能力较差。这两方面的现象都可以归因于区块链为保证链上数据的安全性所引入的拜占庭容错共识协议。GriDB巧妙地将两种方式相结合,将链下扩展的思想运用到跨分片事务处理中,解决了跨分片事务处理慢这一难题。

3　区块链数据可信性

在区块链系统中,数据的存储与查询均由网络中的节点执行并响应用户的查询请求。由于区块链网络中的各节点之间并不互相信任,存储在其他节点上的数据可能被恶意篡改或丢失,为此,需要借助可验证查询技术对不可信节点返回的数据进行验证,保障数据的可信性。

3.1　基于ADS的数据可信保障技术

可验证数据结构(ADS)[36]是一种区块链数据的共识和验证方式,通过计算数据的摘要信息

表1　区块链存储技术的对比

Table 1　Comparison of blockchain storage techniques

系统名称	许可类型	存储模型	扩展方案	跨分片事务	TPS
SEBDB[12-13]	许可链	关系型	链下扩展	–	10^2~10^3
BRD[14]	许可链	关系型	–	–	10^2~10^3
FalconDB[15]	许可链	关系型	–	–	10^2~10^4
Elastico[25]	无许可链	键-值型	分片	不支持	N/A
OmniLedger[26]	无许可链	键-值型	分片	支持	10^4~10^6
Monoxide[27]	无许可链	键-值型	分片	支持	10^4
BrokerChain[28]	无许可链	键-值型	分片	支持	10^3
SharPer[32]	许可链	键-值型	分片	支持	10^4
Pyramid[33]	许可链	键-值型	分片	支持	10^3~10^4
Meepo[34]	许可链	键-值型	分片	支持	10^4~10^5
GriDB[35]	许可链	关系型	分片	支持	10^3

并生成相应证明,保证了区块链上数据的可信性。本节重点介绍基于ADS的数据可信查询技术。目前在区块链领域,主要有基于哈希和基于密码学累加器等两种实现方案。

3.1.1 基于哈希的ADS

基于哈希的ADS利用了哈希函数的单向性和防碰撞性,可以有效证明集合中存在某一元素,最常见的例子便是Merkle Tree[37]。在Merkle Tree中,每个叶节点对应于一个数据块的哈希,而其他节点则存储子节点哈希值合并成新的字符串的哈希值,以此类推直至计算出根节点。在查询时,除了返回查询的数据外,还会返回该数据所在叶节点到根节点路径上所有兄弟节点的哈希值。利用这些信息,用户可以在本地重新构建Merkle路径并计算Merkle Tree根哈希,将其与区块头中存储的根哈希对比,当两者完全符合时便可以认为数据没有受到篡改。而在Merkle Tree的基础上,衍生出了不同的变体,以支持更多功能,包括Merkle Patricia Trie(MPT)[7]、Merkle Bucket Tree(MBT)[38]和Merkle B-tree[22]等。

MPT在以太坊中被引入,是Merkle Tree和Patricia Trie的组合,从根节点到叶节点的路径表示以太坊账户地址,而叶节点则存储对应账户的信息。在MPT中,节点可以分为分支节点、扩展节点和叶节点。分支节点包含一个17元素数组,每个元素表示账户中的一个十六进制字符。扩展节点和叶节点则分别使用公共前缀和唯一后缀达成压缩路径长度的目的。然而,在以太坊中,每个MPT节点存储在键值型数据库中,并通过其散列值来索引,导致读写放大的问题,影响了MPT的整体性能。

MBT在Merkle Tree的基础上引入了哈希表。数据项将首先被映射到不同的哈希表中,然后哈希表作为Merkle Tree的叶节点参与Merkle Tree根哈希的构建。这种设计为MBT带来了两方面优势:一方面,它利用树形结构,当数据状态改变时,只需要更新路径上节点的哈希值,尽可能降低了重新计算哈希带来的成本;另一方面,利用哈希表维护底层数据可以使得数据均匀分布,避免产生局部热点问题。

Merkle B-tree是在B^+ Tree的基础上引入Merkle Tree的特性来支持外包关系数据库的可信范围查询。具体来说,在构造方面,Merkle B-tree的整体结构与B^+ Tree相同,但其中的每个节点都有一个额外字段记录其子节点的哈希;在验证方面,整体流程则与Merkle Tree中相同。值得注意的是,Merkle B-tree节点直接存储在文件系统中,没有任何后端数据库。因此,Merkle B-tree上的查询比MPT上的查询具有更高的性能。然而,Merkle B-tree的结构对于数据更新顺序敏感[39],因此不适合直接应用于区块链。

3.1.2 基于密码学累加器的ADS

密码学累加器[40]的实现过程和安全假设均与RSA算法相近,通过累加函数将一组元素以抗碰撞的方式映射到乘法群中的一个元素或椭圆曲线上的一个点,使得每一个参加累加值计算的元素都可以证明自身在或不在一个特定的集合中。基于密码学累加器的ADS大小不随集合元素数量的增加而增加,节省了网络带宽,多用于有轻节点设定区块链项目中,典型例子为vChain[41]、vChain+[42]系列。

vChain旨在为含有轻节点的区块链系统增加可信布尔查询的功能,为此,作者基于密码学累加器构建了一种新型ADS,并在区块头中增加AttDigest字段记录每个区块的累加值。具体来说,vChain将区块链上每笔交易通过前缀编码方式转化为一个集合,再根据密码学累加器生成一个累加值,若无符合条件的交易,便可以通过交易数据的累加值向用户证明交易与查询条件不匹配;反之,则直接返回满足查询要求的交易。此外,vChain将数值通过二进制编码转换为集合,根据集合相交与否判断其是否属于给定范围,间接实现了区块链上的可信范围查询。

在vChain的基础上,vChain+采用了更先进的密码学集合累加器,支持语义更加丰富的嵌套集合运算,使得vChain+突破了vChain只支持整型范围的限制,无需编码转换即可支持可信多维查询、浮点型范围查询以及这些类型的组合查询。此外,vChain+引入了滑动窗口技术,避免过时数据带来的额外计算开销。

3.1.3 小结

本节介绍了区块链中两类常见的基于ADS的数据可信查询技术,均可以通过区块链上存储的数据摘要验证数据存储与查询的完整性。两类方案的差异主要体现在时空性能上,基于哈希的ADS方案在证明大小、数据查询时间、证明生成时间、证明验证时间上均达到了$O(\log n)$的复杂度,而基于密码学累加器的ADS虽然有着常数级别的证明大小和证明验证时间,但数据查询时间和证明生成时间达到了$O(\sqrt{n})$的复杂度。具体对比见表2,其中n为数据项数量,b为哈希桶数量。

表2 基于ADS的数据可信查询技术对比

Table 2 Comparison of ADS-based trusted data query techniques

ADS类型	ADS	支持的查询种类	证明大小	证明生成时间	证明验证时间	特点
基于哈希	Merkle Tree[37]	布尔查询	$O(\log n)$	$O(\log n)$	$O(\log n)$	构造简单,适用范围广
	MPT[7]	键值查询	—	—	—	提高访问效率
	MBT[38]	键值查询	$O(\log b)$	$O(\log b)$	$O(\log b)$	避免数据聚集
	Merkle B-tree[22]	范围查询	$O(\log n)$	$O(\log n)$	$O(\log n)$	对更新顺序敏感
基于密码学累加器	vChain[41]	布尔查询 范围查询	$O(1)$	$O(\sqrt{n})$	$O(1)$	查询依赖数据编码
	vChain+[42]	多维查询 范围查询 组合查询	$O(1)$	$O(\sqrt{n})$	$O(1)$	查询能力强,提升查询效率

3.2 基于可信硬件的数据可信保障技术

随着硬件技术的不断发展,可信执行环境(TEE)日趋成熟,为解决区块链数据可信保障问题提供了另一种技术思路。具体来说,TEE通过基于硬件安全的CPU实现了内存隔离,攻击者无法直接读取其中的隐私数据和系统密钥,也无法通过固化的硬件逻辑和硬件层面篡改检测,以此确保相关系统运行过程不被恶意篡改。因此,TEE可在保证计算效率的前提下完成隐私保护的计算。常见的实现方案包括Intel Software Guard Extensions(SGX)[43]和ARM TustZone[44]。

在区块链数据可信查询方面,AuthQX[45-46]借助SGX设计了一种全新的验证方案。在AuthQX中,轻客户端通过可信通道与完整节点连接提出查询请求,而装备了SGX的完整节点则需要SGX的可信内存中执行查询、验证并生成完整性证明。这样的设计有效降低了轻客户端的验证负担,为推广区块链应用起到了积极作用。在具体执行方面,由于当前可信硬件的内存空间有限,作者设计了一种在不可信和可信内存中分层组织数据的机制。即不可信内存中的数据被组织为Merkle B-tree,其中频繁访问的内部节点被缓存在可信内存中作为可信检查点。在可信内存中维护的跳表缓冲新附加的块数据。一旦跳表的容量达到阈值,将通过LRU算法进行更新,从而尽可能减小内存数据转换带来的额外开销。

在区块链事务可信执行方面,Dang等人[47]将TEE与PBFT算法相结合,要求所有节点在发送新消息之前将其摘要存储在可信日志中并经过TEE的签名认证。此举可以防止拜占庭节点向不同节点发出冲突的消息,使得系统可以容忍不超过一半的故障节点,进而提升事务可信执行的能力。此外,作者借助SGX中特定的sgx_read_rand和sgx_get_trusted_time函数实现了可信委员会成员分配,为分片场景下事务可信执行提供新的思路。Fang等人[48]设计了一个基于SGX的两阶段事务执行框架,将事务执行的有关逻辑移入SGX的可信内存中,保证了区块链事务的可信执行,进而可以简化区块链内各节点间的共识过程,从而提高系统整体事务执行效率。SE-Frame[49]同样借助TEE来保障区块链事务预执行的正确性,提高系统的并行性。

综上,以SGX为代表的可信硬件以其硬件层面的安全保障,为区块链的数据完整性提供了一种更加便捷且高效的保护手段。但与此同时也应认识到,当下可信硬件还存在可信内存空间不足等诸多问题制约其在区块链领域大规模应用,需要设计更有效的机制来解决。

3.3 讨论

本节介绍了两种主流的区块链数据可信保障技术。总体而言,基于ADS的相关技术方案成熟且实现成本低廉,广泛应用于主流区块链系统中,但其计算开销较大,对系统性能有一定的影响;而基于可信硬件的方案在近年来逐渐涌现,能否高效利用其有限的可信内存是决定该方案能否进一步发展的重要因素。两种方案各有利弊,需要根据实际应用场景和具体需求选择合适的技术方案。

4 区块链数据可流通性

在区块链技术的发展初期,每个区块链网络都是一个相对独立的生态系统,区块链数据只能在该生态系统中流通。然而,随着区块链技术的不断发展和应用,单一区块链网络无法满足不同场景和需求的应用要求。为了在不同区块链之间实现价值的双向流通,跨链技术应运而生。跨链技术是指将不同的区块链网络连接起来,实现区块链数据的跨链流通和价值交换,从而扩大区块链的应用范围和实用价值。

跨链技术对建立数据可信流通体系具有重大意义。首先,跨链技术可以降低数据流通的成本,使得数据可以更加便捷地在不同的区块链网络之间流通。这可以减少数据转移和转换的时间和费用,降低数据流通的门槛,有助于建立更加高效的数据流通体系;其次,跨链技术可以实现不同区块链网络之间的数据可信共享,有助于促进不同行业、不同企业之间的数据合作与共享,提高数据利用率,促进数据的跨界应用和创新,为数据产业的发展提供有力支撑;最后,跨链技术可以实现不同区块链网络之间的价值流通,有利于构建更加开放、多元化的数据市场。这可以促进数据交易市场的发展,为数据经济的快速发展提供有力保障。

4.1 主流的跨链机制

目前跨链技术的发展还处于早期阶段,为了连接同构或异构的区块链网络,解决区块链之间的数据传输、交易访问等技术难题,学术界和工业界提出了公证人机制、侧链/中继、哈希锁定和混合机制等解决方案。

4.1.1 公证人机制

公证人机制是一种广泛应用且相对容易实现的一种跨链机制。在该机制中,为了使互不信任的两个区块链能够实现价值的流动,最简单的方案是为双方引入了一个共同信任的第三方作为中介,其可以是中心化的机构或者一组节点,被称为公证人。与银行提供的外币兑换服务类似,公证人需要负责跨链数据的收集、交易的确认和验证,以保证跨链交易的可信性和交易执行的原子性。

虽然公证人的引入解决了跨链的互通性问题,但是该机制也带来了公证人不可信的风险。为了有效减少跨链交易对单一公证人的依赖,降低公证人作恶对系统的影响,现有的跨链技术通常采用公证人组的形式处理跨链交易。例如,AgentChain[50]将各个跨链公证人合并为一个交易组,并在现有的区块链上生成一个多重签名地址作为存款池,未经这些公证人多重签名的代币将无法转出存款池。这种基于多重签名的公证人机制既保证了在部分节点故障时不至于影响整个系统的稳定性,还利用公证人组的竞争更好地实现了去中心化。Sun等人[51]还采用基于信誉值的选举办法,从拥有较高信誉值的候选公证人中随机选取公证人作为处理跨链交易的中介,并动态更新信誉值以限制公证人作恶。

4.1.2 侧链/中继

侧链机制的原理是在主链的基础上开辟一条新的侧链,将主链上的资产转移到侧链上进行交易和监管,最终再将结果返回给主链。作为一个独立的区块链,侧链通常可以拥有自己的独立

协议、共识机制等,可以在不干扰主链的情况下实现高效的数据管理。侧链机制的优势在于它能够扩展主链的容量和功能,并支持主链和侧链的数据流通,用户可以方便地将主链上的资产或其他数据要素通过跨链交易转移到侧链上进行各种应用的操作,操作完成后再将其转回主链。

中继机制的原理是在不同的区块链之间构建一个中继网络,将交易和数据在中继网络上传递,从而实现不同区块链之间的互操作性。中继网络既可以通过智能合约来管理不同区块链之间的交互,还可以拥有独立的代币和共识机制,此时中继网络便是一个独立的区块链,也称为中继链。中继链相当于从参与跨链的各主链中分离出的上层区块链,在维持各个主链的独立性和安全性的同时支持数据要素在多链间的流通。

常见的侧链/中继机制的应用有BTC Relay[52]和Cosmos[53]。BTC Relay是一种基于以太坊智能合约的中继机制,旨在将比特币网络的信息引入以太坊智能合约中,从而在以太坊上构建去中心化应用。它利用了比特币的SPV协议和Merkle Tree,通过以太坊智能合约来验证比特币交易的状态,实现了跨链价值传输。Cosmos也是一种中继机制,其采用了一种称为"区域"的结构,其中不同区域的区块链之间可以通过中继网络进行交互,实现跨链价值传输和数据交换。

4.1.3　哈希锁定

哈希锁定是一种基于哈希函数和时间锁定的跨链技术,它可以在无需公证人的情况下实现安全可靠的跨链交易,被广泛应用于比特币[5]等区块链之间的跨链交易中。

哈希锁定的原理是利用哈希函数生成一段特定的哈希值作为交易的条件。交易的发起方需要提供这个哈希值,并设置一个过期时间。接收方在收到交易后,需要提供一个满足哈希条件的原像(即原始数据),才能解锁交易并进行后续操作。如果接收方无法提供正确的原像,那么交易将会自动取消,发起方也可以取回自己的资产。

哈希锁定的优势在于具有高度的安全性和灵活性。因为哈希值是无法逆推的,所以即使交易内容被泄露,攻击者也无法窃取资产。同时,哈希锁定还支持跨链多次转账,可以在不同的区块链之间完成多次交易,使得跨链操作更加灵活。

4.1.4　混合机制

目前已有的主流跨链机制在不同方面拥有优势和不足。例如公证人机制实现容易且简洁,但面临中心化的风险;侧链/中继机制可以对区块链扩容,但交易速度慢;哈希锁定只适用于资产交换的场景等。为弥补单个跨链机制的不足,研究者将不同跨链技术相结合,旨在通过混合机制来保证跨链交互的高效和容错性,为用户提供更安全且功能齐全的跨链服务协议。例如,Sun等人[51]基于公证人机制和哈希锁定设计了一个去中心化的跨链协议,通过引入公证人组来改进公证人方案,并设置奖惩和惩罚以控制哈希锁定中交易的解锁时间,确保协议能够抵抗恶意节点的失败者攻击和虫洞攻击,实验证明,这种混合机制在容错性和交易的时间成本上优于原有的单个方案。

4.2　跨链数据流通的难题

虽然上述跨链机制都在不同程度上解决了区块链之间数据隔离和互操作性的问题,但随着区块链应用场景的不断拓展和深入,在跨区块链的数据流通场景中,跨链技术仍存在一些难题亟待解决,例如跨链数据的隐私保护、跨链数据的高效共享等。

在数据要素的隐私保护方面,国务院在《关于构建数据基础制度更好发挥数据要素作用的意见》一文中明确指出:要创新技术手段,推动个人信息匿名化处理等技术的发展及应用,重视个人信息数据在流通和使用时带来的信息安全和个人隐私等问题[4]。在跨链数据的隐私性方面,现有的跨链协议很少考虑对跨链交易信息和参与方身份的隐私保护,这些信息在大部分跨链交易的方案中是公开的[54]。因此,跨链交易环节暴露出的隐私安全问题可能会进一步损害双方区块链的安全性和可用性,对区块链间的数据合作和共享带来不利影响。

针对跨链数据的隐私保护,目前一种有效的解决方案是使用密码学的手段,例如零知识证明等技术。为防范远程侧信道攻击对跨链交换的影响,Li等人[55]利用累加器机制和一种名为CP-SNARK的零知识证明协议在确认块中证明交易,而不会透露交易细节,结果证明该协议实现了跨链交易的公平性、保密性和不可链接性。Xie等人[56]提出现有的跨链桥解决方案要么存在性能问题,要么依赖于会显著降低安全性的委员会。为此他们提出了一种高效的跨链桥方案zkBridge,其应用了非交互式的零知识证明技术zk-SNARK,通过分布式的证明生成协议和简洁的、固定大小的证明压缩方案,降低了跨链交易的验证成本并保证了交易执行的正确性。

在跨链数据的高效共享场景中,数据的传输和验证需要时间和成本,例如PoW侧链[57]的跨链证明大小依赖于链的长度,因此带宽和存储开销会随着区块链的增长而变大,而PoS侧链[58]尽管实现了简洁的证明,但投票周期较长,无法满足快速的跨链转账。基于PoW侧链和PoS侧链,Yin等人[59]采用了公证人+侧链的混合机制,将公证人委员会的选举从共识过程中解耦出来,并采用一种新颖的利用哈希函数产生固定规模的委员会选举方法,降低了投票阶段的总延迟,同时利用委员会生成的跨链证书降低跨链交易验证的成本,从而提高跨链传输的及时性并减少带宽和存储开销。此外,跨链数据共享还面临着查询低效率的不足和结果不一致的风险,为此Cao等人[60]将区块链分为以公有链为基础的授权链和以联盟链为基础的访问链,利用授权链记录数据注册和权限变更信息,并将数据访问控制策略和访问事件记录在不同信任域的访问链中,通过依次追溯数据注册交易、权限变更交易、策略创建交易和访问记录交易,可以重构跨域数据交易、权限授予、访问流程的全周期追溯结果。类似的方案有Vassago[61],其设计了一个两层存储结构:底层是参与跨链的区块链,又叫事务区块链;上层是依赖区块链,负责记录下层的跨链事务依赖。基于构建的跨链依赖图,Vassago在保证查询结果的一致性的前提下,可以在多链上并行化查询,提高查询效率。

5　区块链数据可追溯性

数据溯源是对数据源头及其演变过程的追踪与描述[62]。由于区块链去中心化地保存了系统运行过程中的全部历史交易记录,可溯源成为区块链系统最主要的应用特性之一。目前,包括供应链[63-64]、科学计算[65]、原创成果保护[66]、网络舆情治理[67-68]等多种领域都成功应用区块链技术实现数据来源追溯。

然而,区块链系统以时间顺序将历史交易数据记录在区块链账本上的方式只是为历史数据回溯提供了数据支持,没有提供高效的数据溯源查询能力。现有溯源系统主要采用以下两种方式进行历史数据检索:

(1)基于区块链的数据溯源:由于区块链没有提供溯源查询功能,数据溯源只能通过顺序遍历从创世区块到当前区块记录的交易数据或状态数据信息来完成。这种方式需要花费大量时间,随着系统持续运行,账本数据逐渐增多,数据溯源的代价越来越高。因此,这种方式难以应用在大数据系统中。

(2)基于链下数据库的数据溯源:将区块链数据以溯源查询所需的格式存储在链下数据库中可以有效解决历史数据查询效率问题。在这种方式中,尽管交易记录和状态数据记录可以到区块链中进行真实性验证,但数据演变过程的真实性验证需要重放历史状态所涉及的全部区块交易。

为提供溯源查询的原生支持,部分工作通过设计新型默克尔索引结构来实现支持溯源查询的高效验证。Watanabe等人[69]提出一种新型代币设计,将产品来源和分销过程映射到区块链上,并基于Merkle DAG(Directed Acyclic Graph)对代币状态转换进行建模,通过提交、合并、拆分和分叉等Git语义对供应链中商品上链、加工、分销等过程进行描述,从而记录区块链上代币合约的用户操作轨迹,以实现供应链场景下的数据溯源查询能力。LineageChain[70-71]重新定义了区块

链状态数据节点,并将状态数据以Merkle DAG的形式组成有向无环图,图的边上记录了促使状态改变的交易信息,以增强区块链的存储层,并提供高效的跟踪和篡改证据。进一步地,它用一个确定性的只追加跳表(DASL)来索引Merkle DAG中的节点,来加速溯源查询过程中对DAG的搜索,同时支持前向溯源。DASL利用了区块链的只追加和非随机属性,以区别于普通的跳转列表。这种方案实现了快速和低成本的历史数据查询,使得在智能合约中进行在线溯源成为可能。

在数据查找方面,Wu等人[63]通过研究溯源查询的并行处理来提高溯源查询效率,该工作将记录复制到多个块中,采用并行搜索策略来提高时间效率。为提高搜索的并行性,本文将记录和区块建模为二分图,并使用二分图的最大匹配算法解决分配问题。实验结果表明,在可承受的存储开销下,时间开销最多可减少85.1%。

此外,为保证系统所记录的溯源信息的准确性,如何进行溯源数据的记录与构建也是亟需解决的问题。Cui等人[72]对供应链管理过程中的系统操作流程进行进一步设计,全面解决运输中盗窃、人为错误、交付和管理失败以及供应链中不诚实实体所带来的隐患。其提出了一种基于两笔交易的所有权管理方法。发送方发送所有权转让交易后,需要接收方进行额外的交易确认。一旦发送方和接收方达成协议,则所有权转让将完成。这种方式可以自动标记在运输过程中的物品丢失、人为错误和交付失败等信息。

目前,区块链可追溯性方面的研究主要集中在如何利用区块链系统原有特性实现各个不同领域的数据溯源,而如何改进区块链系统设计使其支持更加完备、高效的数据溯源需求还有待进一步探究。

6　结论与展望

本文从区块链数据的可用性、可信性、可流通性、可追溯性4个方面,探讨了基于区块链的数据要素可信流通关键技术。在区块链数据可用性方面,越来越多的研究尝试在区块链上建立形式更加丰富的存储模型,以适应不同种类数据要素存储需求,为基于区块链的数据要素流通打下了良好基础;而受限于区块链的安全保障机制,目前区块链系统的可扩展性普遍不如传统商业数据库,一定程度上制约了区块链技术的进一步发展和应用。在区块链数据可信性方面,最近的研究聚焦于利用轻量级的可验证结构(ADS)代替传统区块链中基于投票的数据验证方式,提升了数据验证效率,接下来需要进一步如何更好地平衡证明生成时间、证明大小、验证时间等方面的开销;而可信硬件的发展则为这一方向的研究带来了新的思路和挑战。在区块链数据可流通性方面,虽然现有跨链技术已显著促进了数据要素在不同区块链之间的流通,但目前跨链技术对于多区块链的兼容性还难以满足数据要素高效流通的需求。整体而言,跨链技术仍处于起步探索阶段,尚未形成成熟的体系架构。在区块链数据可追溯性方面,区块链系统溯源工作仍处于应用探索阶段,如何基于区块链设计更加完备、高效的数据溯源管理系统还有待研究。

此外,为构建完备的数据要素可信流通体系,区块链技术尚有许多值得深入探讨和研究的问题。本文最后列举一些可能的研究方向,希望对本领域的其他研究者有所启发:

(1)在数据隐私方面:区块链促进了数据流通,但同时造成了数据隐私泄露的问题。现有研究工作聚焦于使用密码学技术实现私有数据的访问控制[73],将敏感数据存储在更好地支持数据隐私的链下数据库中[12]。然而,如何构建数据分级分类管理体系,实现细粒度的数据共享流通仍面临挑战。

(2)在数据使用方面:由于数据要素具有可复制性、非消耗性、边际成本零等不同于传统生产要素的特点,为了保护数据所有权,其在使用过程中往往需要遵循"原始数据不出域、数据可用不可见"的要求[4],影响了数据使用的自由性。现有研究工作将联邦学习与区块链技术相结合[74]以解决上述问题。然而,联邦学习涉及多个参与方之间的频繁通信,区块链拜占庭共识条件下的

通信开销和延迟会影响联邦学习的效率和性能。

(3)在数据表示方面:互联网应用的发展催生了图数据、文档数据等更加多样的数据形式,需要区块链的进一步支持。以图数据为例,针对图的挖掘和分析可以从大规模图数据中提取有用的信息和见解。通过将图数据库与区块链集成,可以以安全和去中心化的方式分析事务关系[75]。

利益冲突声明

所有作者声明不存在利益冲突关系。

参考文献

[1] 国家互联网信息办公室. 数字中国发展报告(2021年)[R]. 2022.

[2] 数据要素白皮书(2022年). 数据要素白皮书(2022年)[R]. 2023.

[3] 中国信息通信研究院安全研究所. 数据要素流通视角下数据安全保障研究报告(2022年)[R]. 2022.

[4] 中共中央国务院. 关于构建数据基础制度更好发挥数据要素作用的意见[EB/OL]. [2023-04-18]. http://www.gov.cn/zhengce/2022-12/19/content_5732695.htm.

[5] 张志威, 王国仁, 徐建良, 等. 区块链的数据管理技术综述[J]. 软件学报, 2020, 31(9): 2903-2925.

[6] NAKAMOTO S. Bitcoin: A peer-to-peer electronic cash system[R/OL]. (2008). https://bitcoin.org/bitcoin.pdf.

[7] BUTERIN V. Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform[R/OL]. (2014)[2023-05-03]. https://ethereum.org/669c9e2e2027310b6b3cdce6e1c52962/Ethereum_Whitepaper_-_Buterin_2014.pdf.

[8] ANDROULAKI E, BARGER A, BORTNIKOV V, et al. Hyperledger fabric: a distributed operating system for permissioned blockchains[C]//Proceedings of the Thirteenth EuroSys Conference. Association for Computing Machinery, 2018: 1-15.

[9] ZHU C, LI J, ZHONG Z, et al. A Survey on the Integration of Blockchains and Databases[J]. Data Science and Engineering, 2023, 8(2): 196-219.

[10] 区块链+数字经济发展白皮书(上)[N]. 中国计算机报, 2021-11-22(008).

[11] AntChain[EB/OL]. https://antchain.net/.

[12] ZHU Y, ZHANG Z, JIN C, et al. SEBDB: Semantics Empowered BlockChain DataBase[C]//2019 IEEE 35th International Conference on Data Engineering (ICDE). 2019: 1820-1831.

[13] ZHU Y, ZHANG Z, JIN C, et al. Towards Rich Qery Blockchain Database[C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. Association for Computing Machinery, 2020: 3497-3500.

[14] NATHAN S, GOVINDARAJAN C, SARAF A, et al. Blockchain meets database: design and implementation of a blockchain relational database[J]. Proceedings of the VLDB Endowment, 2019, 12(11): 1539-1552.

[15] PENG Y, DU M, LI F, et al. FalconDB: Blockchain-based Collaborative Database[C]//Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. Association for Computing Machinery, 2020: 637-652.

[16] THE POSTGRESQL GLOBAL DEVELOPMENT GROUP. PostgreSQL: The World's Most Advanced Open Source Relational Database[EB/OL]. https://www.postgresql.org/.

[17] ZHANG Y, KATZ J, PAPAMANTHOU C. IntegriDB: Verifiable SQL for Outsourced Databases[C]//Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. Association for Computing Machinery, 2015: 1480-1491.

[18] MySQL[EB/OL]. https://www.mysql.com/.

[19] LIU Y, LIU J, VAZ SALLES M A, et al. Building blocks of sharding blockchain systems: Concepts, approaches, and open problems[J]. Computer Science Review, 2022, 46: 100513.

[20] ZHANG C, XU C, XU J, et al. GEM^2-Tree: A Gas-Efficient Structure for Authenticated Range Queries in Blockchain[C]//2019 IEEE 35th International Conference on Data Engineering (ICDE). 2019: 842-853.

[21] ZHANG C, XU C, WANG H, et al. Authenticated Keyword Search in Scalable Hybrid-Storage Blockchains[C]//2021 IEEE 37th International Conference on Data Engineering (ICDE). 2021: 996-1007.

[22] LI F, HADJIELEFTHERIOU M, KOLLIOS G, et al.

Dynamic authenticated index structures for outsourced databases[C]//Proceedings of the 2006 ACM SIGMOD international conference on Management of data. Association for Computing Machinery, 2006: 121-132.

[23] HASSANZADEH- NAZARABADI Y, KÜPÇÜ A, ÖZKASAP Ö. LightChain: Scalable DHT- Based Blockchain[J]. IEEE Transactions on Parallel and Distributed Systems, 2021, 32(10): 2582-2593.

[24] 黄华威, 孔伟, 彭肖文, 等. 区块链分片技术综述[J]. 计算机工程, 2022, 48(6): 1-10.

[25] LUU L, NARAYANAN V, ZHENG C, et al. A Secure Sharding Protocol For Open Blockchains[C]// Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. Association for Computing Machinery, 2016: 17-30.

[26] KOKORIS- KOGIAS E, JOVANOVIC P, GASSER L, et al. OmniLedger: A Secure, Scale- Out, Decentralized Ledger via Sharding[C]//IEEE Symposium on Security and Privacy. IEEE Computer Society, 2018: 583-598.

[27] WANG J, WANG H. Monoxide: Scale out Blockchains with Asynchronous Consensus Zones[C]//16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19). 2019: 95-112.

[28] HUANG H, PENG X, ZHAN J, et al. BrokerChain: A Cross-Shard Blockchain Protocol for Account/Balance- based State Sharding[C]//IEEE INFOCOM 2022 - IEEE Conference on Computer Communications. 2022: 1968-1977.

[29] ZAMANI M, MOVAHEDI M, RAYKOVA M. RapidChain: Scaling Blockchain via Full Sharding[C]//Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. Association for Computing Machinery, 2018: 931-948.

[30] CHEN H, WANG Y. SSChain: A full sharding protocol for public blockchain without data migration overhead[J]. Pervasive and Mobile Computing, 2019, 59: 101055.

[31] AMIRI M J, AGRAWAL D, EL ABBADI A. On Sharding Permissioned Blockchains[C]//2019 IEEE International Conference on Blockchain (Blockchain). 2019: 282-285.

[32] AMIRI M J, AGRAWAL D, EL ABBADI A. SharPer: Sharding Permissioned Blockchains Over Network Clusters[C]//Proceedings of the 2021 International Conference on Management of Data. Association for Computing Machinery, 2021: 76-88.

[33] HONG Z, GUO S, LI P, et al. Pyramid: A Layered Sharding Blockchain System[C]//IEEE INFOCOM 2021 - IEEE Conference on Computer Communications. 2021: 1-10.

[34] ZHENG P, XU Q, ZHENG Z, et al. Meepo: Sharded Consortium Blockchain[C]//2021 IEEE 37th International Conference on Data Engineering (ICDE). 2021: 1847-1852.

[35] HONG Z, GUO S, ZHOU E, et al. GriDB: Scaling Blockchain Database via Sharding and Off- Chain Cross- Shard Mechanism[J]. Proceedings of the VLDB Endowment, 2023, 16(7): 1685-1698.

[36] TAMASSIA R. Authenticated Data Structures[C]//Algorithms- ESA 2003. Berlin, Heidelberg: Springer, 2003: 2-5.

[37] PANG H, JAIN A, RAMAMRITHAM K, et al. Verifying completeness of relational query results in data publishing[C]//Proceedings of the 2005 ACM SIGMOD international conference on Management of data. Association for Computing Machinery, 2005: 407-418.

[38] YUE C, XIE Z, ZHANG M, et al. Analysis of Indexing Structures for Immutable Data[C]//Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. Association for Computing Machinery, 2020: 925-935.

[39] WANG S, DINH T T A, LIN Q, et al. Forkbase: an efficient storage engine for blockchain and forkable applications[J]. Proceedings of the VLDB Endowment, 2018, 11(10): 1137-1150.

[40] BENALOH J, DE MARE M. One- Way Accumulators: A Decentralized Alternative to Digital Signatures [C]//Advances in Cryptology—EUROCRYPT'93. Berlin, Heidelberg: Springer, 1994: 274-285.

[41] XU C, ZHANG C, XU J. vChain: Enabling Verifiable Boolean Range Queries over Blockchain Databases[C]//Proceedings of the 2019 International Conference on Management of Data. Association for

Computing Machinery, 2019: 141-158.

[42] WANG H, XU C, ZHANG C, et al. vChain+: Optimizing Verifiable Blockchain Boolean Range Queries [C]//2022 IEEE 38th International Conference on Data Engineering (ICDE). 2022: 1927-1940.

[43] KARANDE V, BAUMAN E, LIN Z, et al. SGX-Log: Securing System Logs With SGX[C]//Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security. Association for Computing Machinery, 2017: 19-30.

[44] SANTOS N, RAJ H, SAROIU S, et al. Using ARM trustzone to build a trusted language runtime for mobile applications[C]//Proceedings of the 19th international conference on Architectural support for programming languages and operating systems. Association for Computing Machinery, 2014: 67-80.

[45] SHAO Q, PANG S, ZHANG Z, et al. Authenticated Range Query Using SGX for Blockchain Light Clients[C]//Database Systems for Advanced Applications. Cham: Springer International Publishing, 2020: 306-321.

[46] PANG S, SHAO Q, ZHANG Z, et al. AuthQX: Enabling Authenticated Query over Blockchain via Intel SGX[C]//Database Systems for Advanced Applications. Cham: Springer International Publishing, 2020: 727-731.

[47] DANG H, DINH T T A, LOGHIN D, et al. Towards Scaling Blockchain Systems via Sharding[C]//Proceedings of the 2019 International Conference on Management of Data. Association for Computing Machinery, 2019: 123-140.

[48] FANG M, ZHANG Z, JIN C, et al. High-Performance Smart Contracts Concurrent Execution for Permissioned Blockchain Using SGX[C]//2021 IEEE 37th International Conference on Data Engineering (ICDE). 2021: 1907-1912.

[49] FANG M, ZHOU X, ZHANG Z, et al. SEFrame: An SGX-enhanced Smart Contract Execution Framework for Permissioned Blockchain[C]//2022 IEEE 38th International Conference on Data Engineering (ICDE). 2022: 3166-3169.

[50] LI D, LIU J, TANG Z, et al. AgentChain: A Decentralized Cross-Chain Exchange System[C]//2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE). 2019: 491-498.

[51] SUN Y, YI L, DUAN L, et al. A Decentralized Cross-Chain Service Protocol based on Notary Schemes and Hash-Locking[C]//2022 IEEE International Conference on Services Computing (SCC). 2022: 152-157.

[52] BTC Relay[EB/OL]. [2023-05-03]. https://github.com/ethereum/btcrelay/wiki.

[53] KWON J, ETHAN B. A network of distributed ledgers cosmos[EB/OL]. https://v1.cosmos.network/intro.

[54] 孟博, 王乙丙, 赵璨, 等. 区块链跨链协议综述[J]. 计算机科学与探索, 2022, 16(10): 2177-2192.

[55] LI Y, WENG J, LI M, et al. ZeroCross: A sidechain-based privacy-preserving Cross-chain solution for Monero[J]. Journal of Parallel and Distributed Computing, 2022, 169: 301-316.

[56] XIE T, ZHANG J, CHENG Z, et al. zkBridge: Trustless Cross-chain Bridges Made Practical[C]//Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. Association for Computing Machinery, 2022: 3003-3017.

[57] KIAYIAS A, ZINDROS D. Proof-of-Work Sidechains [C]//Financial Cryptography and Data Security. Cham: Springer International Publishing, 2020: 21-34.

[58] GAŽI P, KIAYIAS A, ZINDROS D. Proof-of-Stake Sidechains[C]//2019 IEEE Symposium on Security and Privacy (SP). 2019: 139-156.

[59] YIN L, XU J, TANG Q. Sidechains With Fast Cross-Chain Transfers[J]. IEEE Transactions on Dependable and Secure Computing, 2022, 19(6): 3925-3940.

[60] CAO L, ZHAO S, GAO Z, et al. Cross-chain data traceability mechanism for cross-domain access[J]. The Journal of Supercomputing, 2023, 79(5): 4944-4961.

[61] HAN R, XIAO J, DAI X, et al. Vassago: Efficient and Authenticated Provenance Query on Multiple Blockchains[C]//2021 40th International Symposium

on Reliable Distributed Systems (SRDS). 2021: 132-142.

[62] 刘海鸥, 何旭涛, 李凯, 等. 区块链数据溯源机制研究综述[J]. 情报杂志, 2022, 41(7): 100-106.

[63] WU H, JIANG S, CAO J. High-Efficiency Blockchain-Based Supply Chain Traceability[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(4): 3748-3758.

[64] ORJUELA K G, GAONA-GARCÍA P A, MARIN C E M. Towards an agriculture solution for product supply chain using blockchain: case study Agro-chain with BigchainDB[J]. Acta Agriculturae Scandinavica, Section B — Soil & Plant Science, 2021, 71(1): 1-16.

[65] AL-MAMUN A, YAN F, ZHAO D. SciChain: Blockchain-enabled Lightweight and Efficient Data Provenance for Reproducible Scientific Computing[C]// 2021 IEEE 37th International Conference on Data Engineering (ICDE). 2021: 1853-1858.

[66] ZHU P, HU J, LI X, et al. Using Blockchain Technology to Enhance the Traceability of Original Achievements[J]. IEEE Transactions on Engineering Management, 2023, 70(5): 1693-1707.

[67] DE SOTO H. A tale of two civilizations in the era of Facebook and blockchain[J]. Small Business Economics, 2017, 49(4): 729-739.

[68] HUCKLE S, WHITE M. Fake News: A Technological Approach to Proving the Origins of Content, Using Blockchains[J]. Big Data, 2017, 5(4): 356-371.

[69] WATANABE H, ISHIDA T, OHASHI S, et al. Enhancing Blockchain Traceability with DAG-Based Tokens[C]//2019 IEEE International Conference on Blockchain (Blockchain). 2019: 220-227.

[70] RUAN P, CHEN G, DINH T T A, et al. Fine-grained, secure and efficient data provenance on blockchain systems[J]. Proceedings of the VLDB Endowment, 2019, 12(9): 975-988.

[71] RUAN P, DINH T T A, LIN Q, et al. LineageChain: a fine-grained, secure and efficient data provenance system for blockchains[J]. The VLDB Journal, 2021, 30(1): 3-24.

[72] CUI P, DIXON J, GUIN U, et al. A Blockchain-Based Framework for Supply Chain Provenance[J]. IEEE Access, 2019, 7: 157113-157125.

[73] RUAN P, KANZA Y, OOI B C, et al. LedgerView: Access-Control Views on Hyperledger Fabric[C]// Proceedings of the 2022 International Conference on Management of Data. Association for Computing Machinery, 2022: 2218-2231.

[74] CAO M, ZHANG L, CAO B. Toward On-Device Federated Learning: A Direct Acyclic Graph-Based Blockchain Approach[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(4): 2028-2042.

[75] TSOULIAS K, PALAIOKRASSAS G, FRAGKOS G, et al. A Graph Model Based Blockchain Implementation for Increasing Performance and Security in Decentralized Ledger Systems[J]. IEEE Access, 2020, 8: 130952-130965.

收稿日期:2023年5月4日

钟子岳,北京理工大学计算机学院,博士研究生,主要研究方向为区块链系统。

本文主要承担工作为论文选题,全文统筹规划,论文主要部分的撰写与修改。

ZHONG Ziyue is a Ph.D. student at the School of Computer Science & Technology, Beijing Institute of Technology. His research interests include blockchain systems.

In this paper, he is responsible for selecting the research topic, coordinating the overall planning of the paper, and writing and revising the main sections of the paper.

E-mail: ziyue_zhong@bit.edu.cn

张美慧,北京理工大学计算机学院,教授,曾于2014年至2017年任新加坡科技设计大学助理教授。研究方向包括数据分析、海量数据集成、分布式和区块链系统。担任VLDB'18、IEEE ICDE'18、VLDB'19、VLDB'20、SIGMOD'21和ICDE'22的程序委员会副主席/副主编。

本文中负责选题与写作指导、论文最终审定。

ZHANG Meihui is currently a professor with the Beijing

Institute of Technology. She was an assistant professor with the Singapore University of Technology and Design (SUTD), from 2014 to 2017. Her research interests include data analytics, massive data integration, distributed and blockchain systems. She has served as a vice PC chair/associate editor for VLDB'18, IEEE ICDE'18, VLDB'19, VLDB'20, SIGMOD'21, and ICDE'22.

In this paper, she is responsible for topic and writing instruction, paper review and editing.

E-mail: meihui_zhang@bit.edu.cn

引文格式: 钟子岳, 朱长昊, 李浚哲, 张美慧. 基于区块链的数据要素可信流通技术综述[J].数据与计算发展前沿, 2023,5(5): 46-62. DOI:10.11871/jfdc.issn.2096-742X.2023.05.005. https://cstr.cn/32002.14.jfdc.CN10-1649/TP.2023.05.005.

ZHONG Ziyue, ZHU Changhao, LI Junzhe, ZHANG Meihui. A Survey on Blockchain-Based Trusted Data Elements Circulation[J]. Frontiers of Data & Computing, 2023, 5(5): 42-62. DOI:10.11871/jfdc.issn.2096-742X.2023.05.005. https://cstr.cn/32002.14.jfdc.CN10-1649/TP.2023.05.005.