区块链数据溯源机制研究综述*

刘海鸥[1,2] 何旭涛[1] 李 凯[1] 高 悦[1]

(1. 燕山大学经济管理学院MBA教育中心 秦皇岛 066004;
2.燕山大学互联网+与产业发展研究中心 秦皇岛 066004)

摘 要:[**研究目的**]梳理区块链数据溯源重要研究成果,为科学评价区块链数据溯源研究效果、促进该领域研究发展与落地应用提供借鉴。[**研究方法**]在阐释区块链及数据溯源概念内涵的基础上,系统梳理区块链数据溯源的典型应用,深入分析区块链数据溯源困境并论述基于区块链的数据溯源机制,最后对区块链数据溯源机制的未来发展方向进行了展望。[**研究结论**]数据溯源信息存储技术亟待更新和完善,数据溯源信息安全管理模型与隐私保护机制尚不规范,研究的创新性与落地实用性有待结合深化。

关键词:区块链;数据溯源;溯源机制;隐私保护;信息存储技术
中图分类号:G353.1;G250.7 **文献标识码**:A **文章编号**:1002-1965(2022)07-0100-07
引用格式: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):100-106,40.
DOI:10.3969/j.issn.1002-1965.2022.07.015

A Literature Review of Blockchain Traceability Mechanism

Liu Haiou[1,2] He Xutao[1] Li Kai[1] Gao Yue[1]

(1. School of Economics and Management, Yanshan University, Qinhuangdao 066004;
2. Internet + and Industrial Development Research Center, Yanshan University, Qinhuangdao 066004)

Abstract: [**Research purpose**] The paper sorts out important research results related to blockchain data traceability, and provides a reference for evaluating the effect of blockchain data traceability research and promoting the research development and application in this field. [**Research method**] On the basis of explaining the connotation of the concept of blockchain and data traceability, the paper systematically combs the typical applications of blockchain data traceability, deeply analyzes the dilemma of blockchain data traceability and discusses the data traceability mechanism based on blockchain, and finally analyzes the blockchain data. The future development direction of the traceability mechanism is also prospected. [**Research conclusion**] Data traceability information storage technology urgently needs to be updated and improved, data traceability information security management models and privacy protection mechanisms are not yet standardized, and the innovation and practicality of research need to be combined and deepened.

Key words: blockchain; data traceability; traceability mechanism; privacy protection; information storage technology

0 引 言

作为继互联网之后的下一代颠覆性核心技术,区块链具有去中心化、分布式存储、数据可溯源等特性,被誉为“下一个信任的基石”。其中,可溯源被视为区块链技术落地的最佳性能之一,许多领域已成功运用区块链技术来实现数据溯源,区块链数据溯源技术可以清楚地找到目标数据的源头,并通过判断目标数据的真实性来维护自己的权利,同时也更加深入地了解目标数据的全面信息等。但在进行数据溯源过程中,如果没有有效的措施来对溯源信息进行保护,在其产生及后续演变过程中也会遭到意外破坏、篡改或删除等。此外,目前有关区块链数据溯源的重要研究成果主要集中在溯源模型、溯源存储以及溯源应用等方面,

收稿日期:2021-08-13 修回日期:2021-11-08
基金项目:国家社会科学基金一般项目“区块链生态赋能的‘个性化推荐—隐私悖论’平衡机制研究”(编号:21BTQ081)研究成果之一。
作者简介:刘海鸥,男,1981年生,博士后,副教授,博士生导师,研究方向:区块链与数据治理;何旭涛,男,1997年生,硕士研究生,研究方向:信息资源管理;李 凯,男,1997年生,硕士研究生,研究方向:数据挖掘;高 悦,女,1997年生,硕士研究生,研究方向:区块链。
通信作者:何旭涛

如 Liu 等[1]构建了一种基于数字水印和区块链的边缘计算分布式溯源模型,这种模型通过划分内外区域和选取主节点,能够提高传统区块链数据溯源模型的安全性,并进一步降低资源受限节点的存储容量,进而实现解决区块链可扩展性差等问题。Amin 等[2]建立了一个支持不可变交易和时间快照的区块链数据溯源信息存储系统,该存储系统通过在关系表中嵌入区块链,使系统数据库以防篡改的方式高效存储数据溯源相关信息。Provenance[3]首次利用区块链进行数据溯源(存储食品从生产到被消费的溯源)。

需要指出的是,如何保障数据溯源的安全性与可靠性是当前学界关注的热点问题,部分学者也针对该问题进行了一定程度的研究,但相关成果还处于碎片化状态,没有形成完备和规范的区块链数据溯源机制研究模型和理论体系。亟需科学梳理区块链数据溯源的概念内涵、应用范围、面临困境及解决机制,开展区块链数据溯源机制专项研究。鉴于此,本文对现有的区块链溯源文献进行了系统梳理,着重介绍目前区块链溯源机制研究领域的最新进展和研究热点,追踪新的研究成果并分析其发展趋势,以期引起国内学者对区块链溯源机制研究领域的关注,为图情学科的信息追溯与信息组织管理提供参考。

1　相关概念内涵

1.1　区块链

中本聪(Satoshi Nakamoto)[4]最早提出了“比特币”概念,其底层技术区块链开始受到关注。此后,学者们试着从不同角度出发对区块链概念进行了定义。如 Melanie[5]在其著作《区块链:新经济蓝图及导读》中指出区块链技术是一种可以全员参与共建、共识的数据库。刘海鸥等[6]把区块链比作一份电子账本,这种电子账本由各个不可修改的单独区块组成,并且是从首页按顺序链接到下一页,各个区块又承担着记录发生在相应区段交易的详细记录。HE 等[7]指出区块链是由共识算法、隐私保护、智能合约等技术组合而成;蔡晓晴等[8]通过研究认为区块链是通过借助去中心化及去信任方式进行集体维护的安全可靠技术方案。

综上所述,在已有的区块链相关概念定义基础上,本文认为区块链在实质上是一种计算机技术在互联网时代的创新应用模式,它有效整合了数据库、密码学以及共识机制等技术,其高度去中心化、集体维护、防伪不可篡改以及溯源可追踪等特征,为数据信用赋能、碎片化服务叠加和场景化服务重置等智慧化应用创造了链式结构。

1.2　数据溯源

数据溯源概念的提出是一个持续完善更新的过程。“数据从哪来”和“有哪些中间数据可用来实现数据溯源”的相关问题促使数据溯源思想开始萌芽;其后,数据世系被提出用于形容目标数据的产生和后续数据的应用转化过程;在后续研究中,Cui 等[9]又进一步扩展了数据溯源的内涵范畴:一是对数据产生直接影响的源头数据,二是发现目标数据在源头数据库中的位置。至此,数据溯源这一术语被正式使用。不同领域对数据溯源的认识存在较大的差别。Goble[10]基于生物信息学视角,认为数据溯源不仅要追踪其源头数据及其演变过程,还需要更加具体的信息来保证目标数据的可重复使用;乔蕊等[11]指出数据溯源在本质上是一种记录目标数据的演变路径及其注释的元数据;Glavic 等[12]指出数据溯源主要有两层含义,一层含义是将目标数据的溯源描述为导致其创建的过程,另一层含义是关注演变数据的原始数据来源。

通过上述分析可以看出,虽然学者对数据溯源定义的侧重点不同,但均重点关注目标数据的源头数据及其后续演变过程。因此,本文认为数据溯源就是针对目标数据的源头数据及其后演变全阶段加以追溯、确认、描述、分析以及最后保存的动态过程。整个过程主要涉及三方面内容:一是对目标数据源头数据的追踪与描述;二是对源头数据怎样演变成当前数据过程的全阶段信息的追溯与记录,具体包括目标数据的移动、演变、执行以及传播和交流等行为,以及在整个过程中产生的派生数据;三是从源头数据到当前数据过程中,对目标数据状态产生影响的各种因素进行追溯、描述以及分析和记录。因此,数据溯源既是从当前数据到源头数据的逆向追溯的过程,也是记录从源头数据到当前数据的整个演变过程。在整个数据溯源过程中,形成了一系列内容丰富、系统科学、紧密联系的数据项集,也即数据溯源的结果信息。

2　区块链数据溯源的典型应用

在系统梳理区块链与数据溯源的相关概念后,本部分从网络舆情、政府数据开放共享、数字图书馆以及农业食品生产等方面进一步分析区块链数据溯源的典型应用。

2.1　网络舆情治理方面的应用

区块链溯源机制在网络舆情领域的应用主要体现为舆情的溯源与管控。如 Soto D[13]借助区块链技术有效解决了网络舆情在传播过程中的信息传播鸿沟问题。Huckle S[14]提出了一种用于追溯数字媒体内容的区块链技术原型,并进一步验证了区块链在数字媒体信息溯源领域的科学有效性。Zhang X 等[15]研究了区块链技术在健康社区舆情信息隐私问题、前因及舆情

信息披露意愿等方面的应用。国内学者黄微等[16]基于扎根理论研究范式,借助区块链技术对舆情用户信息进行溯源,并进一步构建了区块链技术能力对网民舆情信息接受行为的影响模型。赵丹等[17]基于区块链和信息传播理论,提出了区块链环境下的网络舆情信息传播概念模型,基于此概念模型进行网络舆情信息溯源,该研究通过区块链技术破解网络谣言溯源与识别瓶颈,有助于重构网络舆情信息传播生态。

2.2 政府数据开放共享方面的应用

Liang 等[18]提出将政务数据嵌入到区块链来收集和验证云数据的溯源信息,同时保留哈希值以保护隐私不受其他节点的影响。Hofman 等[19]提出创建一个语义法律层来支持基于区块链的法律合同,同时创建一个特定司法管辖区的法律本体以及开发一套保持记录证据特征以及智能合约的方法。戚学详等[20]指出区块链技术是一种去中心化、去信任、可追溯、透明、安全的新兴互联网技术,能有效克服当前政府数据治理存在的问题。董祥千[21]系统阐述了区块链的概念与技术原理,并分析区块链在政府网站信息资源安全保存中的应用价值,然后结合区块链的技术特征,设计基于区块链的政府网站信息资源安全保存流程,最后从网络安全评估、资源加密存储、用户认证授权、信息安全共享等方面提出了具体的保障策略。此外,蔡婷等[22]以无边界管理理论为基础,以区块链技术为技术支撑,从打破智慧政务垂直信息协同边界、智慧政务水平信息协同边界、智慧政务内外信息协同边界与智慧政务地域信息协同边界四种边界入手,分析当前影响无边界化智慧政务推进的阻碍因素,并探索无边界化智慧政务的推进机制问题。

2.3 数字图书馆方面的应用

通过区块链溯源技术,可以对图书馆的数字资源建设、数字资产管理、服务优化管理以及知识产权维护等方面进行深度赋能。如区块链溯源机制有助于为文献资源数据和用户行为数据赋予“生命”,以此优化图书馆服务流程,提高服务对接效率,保证资源传递稳定性,同时还可有效保障数字资源知识产权的唯一性以及对数字资产进行科学管理,进而构建和谐版权生态。同时,基于区块链数据溯源技术的数字图书馆可以为用户提供多样化的数字资产管理服务。区块链数据溯源赋能下的图书馆数字资产管理服务模式通过产品体验与底层资管基础链,可有效解决资管环节中服务流程的去中介化问题与可信透明问题,有助于实现价值互联、信息公开透明以及有迹可循,从而为构建数字图书馆资产管理生态系统提供切实可行的解决方案。也有研究提出构建专属于数字图书馆领域处理流程的 PROV 溯源应用模型,由此促进国家科技成果转化,通过对转化项目各项流程的规范化控制,提高成果数据质量。但值得注意的是,需要长期保存的数字资源在录入到数据存储系统过程中,数据可能会发生一系列变化(例如数据篡改等),因此,如将数据发生的变化记录为溯源信息,就可以在一定程度上保证数字资源的真实性与可靠性。目前已出现支持数据溯源功能的数字资源长期存储系统,如美国一家图书馆自行开发的 DAITSS 系统[23]。

2.4 农业食品生产方面的应用

农产品从生产、加工到最后的分销过程中,涉及到农产品各个方面的信息,而将区块链技术应用到农产品溯源过程中国内外已有相关研究做了介绍。例如,Orjuela 等[24]在数据溯源存在不信任这一问题背景下,提出设计和开发一种基于区块链技术的数据库平台,这种平台主要是为管理农业供应链和控制互联网提出解决方案。Torky 等[25]提出一种新颖的区块链模型,这种模型可用于解决农业精准溯源系统中的一些重大挑战。以此为基础,Demestichas 等[26]概述区块链技术在农产品可追溯领域的具体应用,并就有关区块链集成到可追溯性系统方面的应用进行了广泛的文献综述。国内近几年也有区块链农业溯源的一些具体应用,如学者王志铧等[27]提出一种利用区块链技术的农产品柔性可信溯源解决方案,并建立一种系统模式降低存储结构复杂度以实现可信溯源,以此为基础采用动态追溯机制使系统灵活适应不同生产场景,并将超级账本作为区块链实现方式,对关键数据进行分布式加密存储以提高追溯结果可信性,并以生姜产品为溯源对象,通过剖析产业链上下游产品对应关系确定溯源对象粒度、账本内容与数据格式,对基于区块链的柔性可信溯源系统模型进行验证。

3 区块链数据溯源困境分析

虽然比特币底层支撑技术的区块链项目发展迅速,但区块链数据溯源依旧面临较为严峻的应用困境,如数据溯源存储平台性能缺陷、数据溯源安全危机以及溯源效率瓶颈等一系列问题,相关研究也对此进行了探讨。

3.1 存储性能缺陷

目前基于区块链技术存储数据溯源信息的类型大致有两种。一种是通过比特币公链等公有链进行信息存储,另外一种则是借助更为开源的区块链如比特币、以太坊等构建联盟链进行数据溯源信息存储。但无论是公有链还是基于开源代码的联盟链,都很难避免区块链自身技术性能瓶颈导致的数据溯源信息存储障碍问题,且在大数据时代背景下,无论是政府数据开放共享还是数字图书馆建设方面,都面临着用户或资源数

据规模爆炸式增长这一问题,因此,如何在政府数据溯源或数字图书馆用户(资源)数据溯源过程中对溯源信息进行高效存储,是亟待解决的关键问题。此外,区块链数据溯源系统存储的相关信息大部分来自物联网系统,由于每个处于运输状态的商品都会带有不同的感知设备,在商品持续流通的过程中,上报数据规模会随着物联网感知节点数量的增加而不断增长。同时,为满足物联网海量数据的上报需求,区块链数据溯源信息存储需要强大的信息写入性能,但当前通用的区块链技术如比特币、以太坊等针对大规模信息的写入性能一般,难以满足系统的信息存储需要。国内外学者的相关研究也验证了这一观点,如 Aitzhan 等[28]指出目前基于区块链与数字签名技术的交易信息存储仍面临较大的技术瓶颈;黑一鸣等[29]认为分布式存储可以提高云服务数据信息的存储性能,但信息集中存储于存储服务商,完整性验证需要通过第三方完成,故仍存在一定的存储缺陷。因此,迫切需要探讨更为高效科学的区块链数据溯源信息存储方案,以期更好地应用于数据溯源信息存储管理。

3.2　信息安全危机

一方面,区块链数据溯源的信息安全问题本质上是区块链安全监管问题,例如,如何对农业食品生产过程中食品的安全信息进行溯源是近几年研究的热点问题。值得注意的是,区块链安全监管问题主要涉及两个方面:一是区块链共识机制、密钥管理以及智能合约等自身技术局限带来的安全问题;二是区块链高度去中心化和自治性能特点为现有的网络和数据安全监管技术带来的全新挑战。针对溯源数据信息安全问题,Baker J 等[30]利用区块链作为数据溯源跟踪,将区块链交易用于存储食品从生产到消费者的溯源详细信息;Barber S 等[31]在研究中探讨了以比特币为研究场景的数据溯源系统,提出将研究目标作为编码文件存储在比特币交易数据字段中的技术手段。

另一方面,隐私性与保密性是区块链数据溯源应用能否大力推广的基本条件,这一点在网络舆情治理、政府数据开放共享以及数字图书馆建设显得尤为重要,例如舆情治理数据信息溯源过程中会涉及网络用户的个人隐私,政府数据开放共享过程中会涉及政府部分的相关数据隐私,数字图书馆建设过程中同样会涉及图书馆用户和资源的隐私。因此,如何保护数据溯源信息的隐私是一个值得注意的问题。针对区块链溯源导致的隐私泄露问题,Wang 等[32]采用智能合约与同态加密技术保护用户隐私,相比传统的单一使用智能合约技术来保护用户隐私安全性更高;Lei 等[33]指出零知识证明是一种涉及双方或更多方的协议,对于保护用户交易隐私具有很大的实际价值,可以通过引入零知识证明方法对用户的交易隐私进行保护。

3.3　溯源效率瓶颈

效率是数据溯源过程中非常重要的技术指标。无论是将区块链数据溯源用于网络舆情治理、政府数据开放共享,还是应用于数字图书馆建设与农业食品生产方面,使用方对溯源效率的要求极高,同时还需要对数据溯源质量进行严格把控,而当前的数据溯源尚面临较大的效率瓶颈,对溯源的整体效果造成了较大影响。这一问题也引起了国内外学者的关注,如 Dai 等[34]认为传统的数据溯源方法如顺序溯源法最大的问题就是数据溯源效率不高,数据查询也存在类似的性能瓶颈,并且这种方法还需要较大的数据存储空间。在此基础上,Woodruff 等[35]指出,在顺序溯源过程中加入一定的标注功能有利于提高数据溯源的效率,但存储空间需求大的问题依旧没有得到解决。明华等[36]对已有的数据溯源模型进行了回顾,指出虽然目前的数据溯源模型已颇具规模,但始终存在一个共性问题:模型性能有待提高,数据溯源效率还存在缺陷。因此,溯源效率已成为数据溯源亟需解决的关键问题之一。

4　基于区块链的数据溯源机制

针对区块链数据溯源在实际应用过程中面临的上述困境,部分研究通过建立双链存储机制、安全模型机制以及逆向溯源机制等破解区块链数据溯源面临的诸多难题。

4.1　双链存储机制

双链存储机制以链式结构为基础,凭借链上区块中交易的无序特性构建各项交易的链式结构,进而解决数据溯源过程中的信息存储问题。双链存储机制分为数据溯源信息存储及数据溯源信息查询两部分,具体是利用以太坊交易的附加字段,将父交易的哈希散列作为附加数据添加到区块的交易中,这样才能在数据信息查询时按照链式结构对链上的全部数据进行查询。针对数据溯源过程中存在的数据信息存储性能瓶颈,已有研究从双链存储机制层面给出了具体的解决方案,如 SUN 等[37]在 IPFS 存储环境下实现电子病历的稳定存储和高效共享,构造了一种基于区块链的数据加密存储方案。WANG 等[38]研究了分布式存储系统的数据存储与共享方案,提出了将 IPFS、以太坊和基于属性的加密技术相结合的数据存储架构。为了解决数据难以流通及数据存储问题,刘炜等[39]基于区块链的高度去中心化、不可篡改以及集体维护等性能特点,采用双链结构作为区块链架构,构建了一个双链结构传染病数据共享区块链模型,同时通过 IPFS 获得大容量存储空间,以此解决了区块数据溯源信息存储面

临的空间问题，充分保障了数据溯源信息存储的稳定性与共享的安全性。此外，张利华等[40]为解决高速铁路分布存储的数据溯源信息遭受恶意篡改及存储困难等问题，提出一种基于联盟链的去中心化的双链存储模型，由此安全、可靠、高效地存储铁路沿线监测数据溯源信息。

4.2 安全模型机制

在数据溯源过程中，数据溯源信息易被恶意篡改，导致数据溯源信息面临较大的信息安全隐患。因此，为了保证数据溯源信息的真实性、完整性与可靠性，有必要建立数据溯源安全机制，从多个方面保证数据溯源的信息安全。针对数据溯源的潜在威胁，Hansan等[41]指出凭借安全可信源机制可以对数据溯源信息进行检测，由此判定数据溯源信息是否完整，是否遭到破坏，从而保障其安全性。Zhang等[42]利用改进的数据溯源威胁模型解决了数据库中数据溯源信息管理的两个问题：对每条数据记录进行溯源检查，改变数据溯源信息的序列管理方式。贾大宇等[43]利用已有数据溯源标准及模型对数据溯源安全管理进行了扩展，提出一种分层次的数据溯源安全模型，通过这种数据溯源安全模型来保证数据溯源信息的完整性及可信性。此外，通过分析数据隐私攻击方法可以看出，攻击者主要通过监听数据溯源过程中的网络层信息、交易层信息以及应用层信息来获取数据信息。因此，可以从这三个层面出发进行数据溯源隐私保护。首先，在数据网络层隐私保护方面，有研究提出采用限制接入的方式进行安全防御，该方法可以从根本上加大攻击者对信息网络层进行攻击的难度，但需要对区块链的运行机制进行修改，因此存在一定的局限。目前，该种方法主要应用于私有链或联盟链架构中，如超级账本等[44]。还有研究通过检测和屏蔽的方式对数据隐私进行保护，如Huang等[45]通过行为模式聚类来检测存在缺陷的数据信息，该方法能够有效消除数据隐私安全问题。其次，在数据交易层隐私保护方面，Bitlaunder[46]在研究中提出了基于数据失真技术保护数据隐私的方法；Monero[47]研究了门罗币这一专注于隐私保护的数字货币，设计了基于加密机制的保护方案，以此对数据交易层隐私进行保护。最后，在数据应用层隐私保护方面，有研究提出了具有隐私保护机制的区块链技术，如Meiklejohn等[48]通过找寻比特币地址进行污点分析，由此验证比特币址的身份信息并进行隐私防护；还有学者指出利用冷钱包技术[49]对数据秘钥进行离线缓存，以此防止网络对其进行恶意攻击，保护数据隐私存储介质的安全。

4.3 逆向溯源机制

逆向溯源机制是解决数据溯源效率不高的有效方法，该机制对于目标数据的追踪较为简单，且只需要存储较少的元数据就可以实现有效地追踪，且不需要耗用多余的空间来存储溯源过程中的中间处理信息、溯源全过程的注释信息等，因此在很大程度上可以规避数据存储缺陷问题。针对逆向溯源机制，已有学者进行了相关研究，如Dai等[34]详细论述了数据库中逆置追踪的数据溯源机制，并指出逆向溯源的关键是构造逆向函数，能否构造出有效的逆向函数将直接影响数据溯源查询的效果以及溯源算法的性能，最终决定整个溯源过程的效率；相较于标注法，逆向溯源机制的最大优点是所需存储空间较小。在此基础上，Woodruff等[35]提出了逆置函数反向查询法，该方法通过逆向查询或构造逆向函数对查询求逆，或者说根据转换过程反向推导，由结果追溯至原数据，由于该方法是在必要时才进行计算，因此又叫做Lazzy方法。此外，Xu等[50]简要描述了用于农业食品数据记录和追踪的区块链数据逆向溯源工作原理，并从提高数据透明度、实现数据可追溯、提高食品安全质量监控以及降低交易成本四个方面探讨了区块链数据逆向溯源机制的落地策略。

5 总结与展望

在数据信用深度赋能的时代背景下，区块链数据溯源受到学术界和业界的广泛关注并取得了一定的进展，但仍是一个充满挑战的热点研究领域。此外，虽已建立了相应的数据溯源机制来解决实际应用中的诸多问题，但依旧存在以下待完善之处。

a.数据溯源信息存储技术亟待更新和完善。通过文献梳理可以看出，大数据时代需要充分考虑数据溯源信息的规模及相应数据溯源信息存储系统的存储性能。已有研究大多从增链角度考虑，如采用双链结构[41]作为区块链架构，通过IPFS获得大容量存储空间以此来充分保障数据溯源信息的存储稳定性与共享安全性。在此基础上，有文献提出基于双链存储模型构建联盟链，由此对双链存储模型性能进行进一步优化，该模型能有效提升区块链数据溯源系统的存储性能和数据信息的吞吐量，进而广泛应用于铁路沿线监测数据溯源信息存储[40]。需要指出的是，虽然双链存储机制能在一定程度上改善溯源系统的数据信息存储性能，但由于数据溯源信息不断趋于增长，在数据溯源过程中捕获到的数据溯源信息有可能比原始数据信息本身的规模和体量还要大，这就意味着需要更大的数据信息存储空间以及更低的数据信息存储成本，且存储效率也需要不断增强，因此相应的数据溯源信息存储技术还需要进一步优化和完善。

b.数据溯源信息安全管理模型与隐私保护机制尚

不规范。已有的数据溯源模型及其框架大多是针对具体的指定任务建立,其构建的系统也是在自身系统内对数据溯源信息进行溯源和管理,在访问或获取跨系统的溯源信息时,需要通过现有的 API 接口才能实现,但目前该类接口在扩展性和通用性方面还均有欠缺。因此,如何构建一套通用规范的 API 访问接口就显得尤为必要。同时,现有的数据溯源模型复杂多样,数据溯源信息格式也不尽相同,导致数据溯源信息异构问题突出,进而会影响到数据溯源信息的检索及共享效率。针对此类问题,可通过建立数据溯源威胁模型来规范数据溯源信息访问途径[42]或提出访问控制策略,进而保障数据溯源的信息安全。然而,已有的溯源数据信息访问控制及安全模型机制仍然需要不断优化,如需要持续优化安全模型涉及的参数及变量,以此来改进模型效能,规范溯源信息访问与安全管理机制。此外,区块链自身独有的高度去中心化功能保障了数据溯源的完整性和不可篡改性能,智能合约有效降低了数据溯源的收集成本,提高了数据溯源的应用效率。从文献梳理来看,目前的区块链隐私保护机制有从事前进行防范的方面,如节点准入限制、匿名通信系统的使用以及节点隔离机制的应用等;也有从事后方面的监管措施,如恶意节点的监测和屏蔽等[45]。但不同的保护机制也有其自身的适用条件和一定的场景限制,在实际应用过程中,还需要因地制宜地针对区块链网络的类型和特点制定相应的隐私安全维护方案。

c.区块链数据溯源机制的理论创新性与实践实用性有待结合深化。区块链是目前最为前沿的信息技术之一,其数据溯源机制在很多领域尚未形成有效具体的应用场景和成功经验,甚至已落地的应用和当前技术管理手段与制度产生冲突。但是,越来越多的研究成果充分证明了区块链数据溯源的潜在价值,且随着区块链数据溯源技术的不断成熟,其实际应用场景也会不断扩展,这在一定程度上可进一步验证已有区块链数据溯源的相关标准、模型以及方法,也能够更好地解决现实问题,如在如前所述的网络舆情治理、政府数据开放共享、数字图书馆建设以及农业食品生产等领域,区块链将传统互联网架构提升到集成性创新的全新维度,通过区块链赋能形成“区块链+应用”的各类场景,充分发挥区块链在降本增效方面的巨大潜能,重构各领域的发展模式。此外,在未来的发展中,需要进一步提升政府部门和业界对区块链数据溯源技术的重视程度,业界专家也需针对数据溯源开展加密追踪相关技术的创新研究,在为区块链数据溯源机制与数据隐私保护提供技术支持的同时,协助区块链数据溯源项目的实际应用落地。因此,创新性与实用性的结合将是今后完善区块链数据溯源机制的一个发展方向,需在现有区块链数据溯源技术基础上,不断探索和优化更为合理、系统以及规范的技术准则,为今后大规模的区块链数据溯源应用落地奠定基础。

6　结束语

溯源机制作为区块链的核心特征,可以对数据整个生命周期的演变过程进行实时追踪,从而保障数据溯源信息存储的稳定性与共享的安全性。本文在深入分析区块链数据溯源典型应用及困境的基础上,系统梳理了各种溯源机制的研究成果,并对区块链数据溯源机制的未来发展方向进行了展望。随着 5G 时代的来临,区块链将在新型数字化建设中发挥更加重要的作用,不可避免地会与物联网、云计算、大数据、AI 等新兴技术产生接触。因此,如何深度融合前沿的信息化技术,构建高容量、高效率、规范化的区块链数据溯源机制是一个值得关注的研究方向。

参 考 文 献

[1] Liu Caiyun, Chen Xuehong, Li Jun, et al. A novel data traceability model based on blockchain and digital watermarking in edge computing[J]. Journal of Physics: Conference Series, 2020, 1682(1): 1-9.

[2] Amin Beirami, Ying Zhu, Ken Pu. Trusted relational databases with blockchain: Design and optimization[J]. Procedia Computer Science, 2019, 155: 137-144.

[3] Blockchain: the solution for transparency in product supply chains[EB/OL]. [2020-12-05]. http://www.provenan.org/whitepaper.

[4] Satoshinakamoto. Bitcoin: A peer-to-peer electronic cash system[EB/OL]. [2021-01-20]. https:// bitcoin.org/bitcoin.pdf.

[5] Melanie S. Blockchain: Blueprint for a new economy[M]. O Reilly Media, Inc., 2015.

[6] 黄文娜.移动图书馆用户画像大数据应用的困境与对策——基于区块链理念[J].图书馆学研究,2019(23):26-33.

[7] HeP, Yu G, Zhang Y F, et al. Survey on blockchain technology and its application prospect[J]. Computer Science, 2017, 44(4): 1-7, 15.

[8] 蔡晓晴,邓　尧,张　亮,等.区块链原理及其核心技术[J].计算机学报,2021,44(1):84-131.

[9] Cui Y, Widom J, Wiener J L. Tracing the lineage of view data in a warehousing environment[J]. ACM Transactions on Database Systems(TODS), 2000, 25(2): 179-227.

[10] Goble C. Position statement: musings on provenance, workflow and(Semantic Web) annotations for bioinformatics[C]. Proceedings of Workshop on Data Derivation and Provenance. Chicago, Illinois, USA, 2002: 1-5.

[11] 乔　蕊,曹　琰,王清贤.基于联盟链的物联网动态数据溯源机制[J].软件学报,2019,30(6):1614-1631.

[12] Glavic B, Dittrich K R. Data provenance: a categorization of existing approaches[C]. Datenbanksysteme in Business, Technologie

Und Web(BTW 2007).Aachen,Germany,2007:227-241.
[13] Soto D,Hernando. A tale of two civilizations in the era of facebook and blockchain[J]. Small Business Economics,2017,49(4):729-739.
[14] 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.
[15] Zhang X,Liu S,Chen X,et al. Health information privacy concerns, antecedents, and information disclosure intention in online health communities[J]. Information & Management,2018,55(4):482-493.
[16] 黄 微,李 吉.区块链技术对舆情用户信息接受行为意愿的影响研究[J].情报杂志,2020(10):130-136.
[17] 赵 丹,王晰巍,韩洁平,等.区块链环境下的网络舆情信息传播特征及规律研究[J].情报杂志,2018(9):127-133,105.
[18] Liang X,Shetty S,Tosh D,et al. Provchain:a blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability[C].The 17th IEEE/ACM International Symposium on Cluster,Cloud and Grid Computing (IEEE/ACM CCGrid 2017).Madrid,Spain,2017:468-477.
[19] Hofman D L.Legally speaking:smart contracts,archival bonds, and linked data in the blockchain [C].2017 26th International Conference on Computer Communication and Networks (ICCCN).IEEE,2017:1-4.
[20] 戚学详.区块链技术在政府数据治理中的应用:优势、挑战与对策[J].北京理工大学学报(社会科学版),2018(5):105-111.
[21] 董祥千,郭 兵,沈 艳,等.一种高效安全的去中心化数据共享模型[J].计算机学报,2018,41(5):1021-1036.
[22] 蔡 婷,林 晖,陈武辉,等.区块链赋能的高效物联网数据激励共享方案[J].软件学报,2021,32(4):953-972.
[23] Daitss[EB/OL].[2020-02-09].http://daitss.fcla.edu/.
[24] Orjuela KG,Gaona-Garcia PA,Marin CEM. 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.
[25] Torky M,Hassanein AE.Integrating blockchain and the internet of things in precision agriculture:analysis, opportunities, and challenges[J].Computer And Electronics In Agriculture,2020,178(1):1-15.
[26] Demestichas K, Peppes N, Alexakis T, et al. Blockchain in agriculture traceability systems:a review[J].Applied Sciences,2020,10(12):1-22.
[27] Wang Zhihua,Liu Pingzeng,Song Chenbao,et al. Research on the flexible and credible traceability system of agricultural products based on blockchain[J]. Computer Engineering,2020,46(12):313-320.
[28] Aitzhan N Z,Svetinovic D.Security and privacy in decentralized energy trading through multi signatures,block-chain and anonymous messaging streams[J].IEEE Transactions on Dependable and Secure Computing,2016(99):1-6.
[29] 黑一鸣,刘建伟,张宗洋,等.基于区块链的可公开验证分布式云存储系统[J].信息网络安全,2019(3):52-60.
[30] Baker J,Steiner J.Blockchain:The solution for transparency in product supply chains[J].Provenance:London,UK,2015.
[31] Barber S,Boyen X,Shi E,et al.Bitter to better-how to make bitcoin a better currency[C].International Conference on Financial Cryptography and Data Security. Springer, Berlin, Heidelberg, 2012:399-414.
[32] Wang R,Tsai W T,He J,et al.A medical data sharing platform based on permissioned blockchains[C].International Conference on Blockchain Technology and Application. ACM,2018:12-16.
[33] Lei X,Shah N,Lin C,et al.Enabling the sharing economy:privacy respecting contract based on public blockchain[C].ACM Workshop on Blockchain.ACM,2017.
[34] Dai Chaofan.Theories and approach of data lineage tracing in data warehouse environment [D]. National University of Defence Technology,2002.
[35] Woodruff A,Stonebraker G.Supporting fine-grained data lineage in a database visualization environment[C].Proc of the 13th International Conference on Data Engineering, Washington DC: IEEE Computer Society,1997:91-102.
[36] 明 华,张 勇,符小辉.数据溯源技术综述[J].小型微型计算机系统,2012(9):1-7.
[37] Sun J,Yao X M,Wang S P,et al.Blockchain-based secure storage and access scheme for electronic medical records in IPFS[J].IEEE Access,2020(8):59389-59401.
[38] Wang S P, Zhang Y L.Ablockchain-based framework for data sharing with fine-grained access control in decentralized storage systems[J]. IEEE Access,2018(6):38437-38450.
[39] 刘 炜,李 阳,田 钊,等.IDDS:一种双链结构传染病数据共享区块链模型[J].计算机应用研究,2020(1):1-6.
[40] 张利华,蒋腾飞,姜攀攀,等.基于区块链的高速铁路监测数据安全存储方案[J].计算机工程与设计,2020,41(4):933-938.
[41] Hasan R, Sion R, Winslett M. Introducing secure provenance: problems and challenges[C]. Proceedings of the 2007 ACM Workshop on Storage Security and Survivability (Storage SS). Alexandria,Virginia,USA,2007:13-18.
[42] Zhang J,Chapman A,Lefevre K.Do you know where your data's been? tamper evident database provenance[C].Proceedings of the 6th VLDB Workshop on Secure Data Management. Lyon, France,2009:17-32.
[43] 贾大宇,信俊昌,王之琼,等.存储容量可扩展区块链系统的高效查询模型[J].软件学报,2019,30(9):2655-2670.
[44] Hyperledger.Hyperledger architecture working group paper[EB/OL].[2021-01-10]. http://www. hyperledger.org/.
[45] Huang Butian,Liu Zhenguang,Chen Jianhai,et al.Behavior pattern clustering in blockchain networks [J]. Multimedia Tools&Application,2017,76(19):20099-20110.
[46] BitLaunder. BitLaunder's mixer vs "major exchanges" mixer[EB/OL].[2021-01-20]. http://bitcoin.stack exchange.com/question/25722/BitLaunder's mixer vs "major exchanges" mixer/25753.

(下转第40页)

research/china-as-a-cyber-great-power-beijings-two-voices-in-telecommunications/.
[8] George Perkovich. China-U.S. cyber-nuclear C3 stability[EB/OL].[2021-09-24].https://carnegieendowment.org/2021/04/08/china-u.s.-cyber-nuclear-c3-stability-pub-84182.
[9] 中国政府网.《国家网络空间安全战略》发布[EB/OL].[2021-09-24]. http://www.cac.gov.cn/2016-12/27/c_1120195878.htm.
[10] Adam Segal. An emerging china-centric order China's vision for a new world order in practice[EB/OL].[2021-09-24]. https://www.nbr.org/publication/chinas-vision-for-cyber-sovereignty-and-the-global-governance-of-cyberspace.
[11] Scott W. Harold, Martin C. Libicki, Astrid Stuth Cevallos. Getting to yes with China in cyberspace[EB/OL].[2021-09-24]. https://www.rand.org/pubs/research_reports/RR1335.html.
[12] Jonathan E. Hillman, Maesea McCalpin. Watching Huawei's "safe cities"[EB/OL].[2021-10-12]. https://www.csis.org/analysis/watching-huaweis-safe-cities.
[13] Amy Chang. Warring state China's cybersecurity strategy.[EB/OL].[2021-10-12].https://www.files.ethz.ch/isn/186337/CNAS_WarringState_Chang.pdf.
[14] Samm Sacks & Manyi Kathy Li. How Chinese cybersecurity standards impact doing business in China[EB/OL].[2021-10-12]. https://www.csis.org/analysis/how-chinese-cybersecurity-standards-impact-doing-business-china.
[15] Gregory C. Allen. Understanding China's AI strategy: Clues to Chinese strategic thinking on artificial intelligence and national security[EB/OL].[2021-10-12]. https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy.
[16] Emily de La Bruyère. The network great-power strategy a blueprint for China's digital ambitions[EB/OL].[2021-10-12].https://www.nbr.org/publication/the-network-great-power-strategy-a-blueprint-for-chinas-digital-ambitions.
[17] Benjamin Shobert. China's pursuit of next frontier tech computing, robotics, and biotechnology[EB/OL].[2021-10-12]. https://www.nbr.org/publication/chinas-pursuit-of-next-frontier-tech-computing-robotics-and-biotechnology.
[18] James Andrew Lewis. ZTE, the telecom wars, and cyber spies[EB/OL].[2021-10-12].https://www.csis.org/analysis/zte-telecom-wars-and-cyber-spies.
[19] Ariel Levite. ICT supply Chain integrity: Principles for governmental and corporate policies[EB/OL].[2021-10-12]. https://carnegieendowment.org/2019/10/04/ict-supply-chain-integrity-principles-for-governmental-and-corporate-policies-pub-79974.
[20] Amy Chang. Warring state: China's cybersecurity strategy[EB/OL].[2021-10-12].https://www.cnas.org/publications/reports/warring-state-chinas-cybersecurity-strategy.
[21] William A. Carter. Smart money on Chinese advances in AI[EB/OL].[2021-10-12]. https://www.csis.org/analysis/smart-money-chinese-advance-ai.
[22] Adam Segal. Stabilizing cybersecurity in the U.S.-China relationship[EB/OL].[2021-10-12]. https://www.nbr.org/publication/stabilizing-cybersecurity-in-the-u-s-china-relationship.
[23] 郎 平.全球数字地缘版图初现端倪[J].信息安全与通信保密,2021(3):9-15.
[24] 鲁传颖.网络空间安全困境及治理机制构建[M].2018:55-59.
[25] 徐培喜 .2020 数字冷战元年:网络空间全球治理的两种路线之争[J].信息安全与通信保密 ,2021(3):16-23.
[26] 侯冠华.美国智库对中美科技竞争的观点解读及对策建议[J].情报杂志,2021,40(4):33-41.
[27] 叶圣萱.美国智库对"数字丝绸之路"倡议的认知及启示[J].情报杂志,2021,40(3):70-75,88.
[28] 江天骄.中美网络空间博弈与战略稳定[J].信息安全与通信保密,2020(9):11-17.

(责编/校对:王平军)

(上接第106页)

[47] Monero. A note on chain reactions in traceability in cryptoNote 2.0[EB/OL]. [2021-02-12].https://lab.get-monero.org/pubs/MRL-0001.pdf.
[48] Meiklejohn S, Pomarole M, Jordan G, et al. A fistful of bitcoins: characterizing payments among men with no names[C]//Proc of the 13th ACM Internet Measurement Conf. New York: ACM, 2013:127-140.
[49] Okcoin. OKCoin cold wallet security design and protocol[EB/OL].[2021-01-22]. http://www.okcoin.com/security.html.
[50] Xu Jie, Guo Shuang, Xie David. Blockchain: a new safeguard for agri-foods[J]. Artificial Intelligence in Agriculture, 2020, (4): 153-161.

(责编:王育英;校对:贺小利)