基于区块链技术的海洋数据资源共享应用设计

张驰[1],王瑞[2],程骏超[1],何元安[1],甄君[1]

1. 中国船舶工业系统工程研究院;国防科技工业海洋安全体系创新中心,北京 100036
2. 中国科学院软件研究所,北京 100089

摘要　分析了中国“海洋数据孤岛”问题的本质原因,提出了一种基于区块链技术的海洋数据资源共享应用模式,改良了区块节点的共识机制和海洋数据在区块链上的存储方式,构建,海洋数据资源共享平台,以达到推动海洋数据的开放共享与增值流通的目的。

关键词　海洋数据;区块链;数据共享;增值流通

新中国成立以来,中国持续开展各类海洋调查、观测和监测活动以及国际海洋数据合作与交换工作,获取和积累了一定数量的全学科海洋数据和相关信息,数据范围覆盖中国沿岸、近海乃至全球海域。由于历史条件等原因,早期中国海洋数据多以纸质、光盘或磁带等形式存储,严重限制了数据的共享服务。自20世纪80年代初,陆续开展原有历史数据的数字化工作,同时随着计算机的普及和海洋仪器设备的更新换代,中国海洋数据的使用和共享服务步入信息化时代。尽管目前中国海洋数据共享服务已初步形成了局部或有关部门范围内的业务体系,在科技创新和国家海洋事业发展中发挥了一定作用,但仍存在着一些机制、体系及技术方面的问题,使得国内海洋数据共享和国际交换均存在一定的困难,在一定程度上制约了中国海洋科技和海洋经济等领域的快速可持续发展,更难以满足当前“海洋强国”战略目标建设实施的需求。

1　“海洋数据孤岛”问题成因

1.1　海洋数据规范与标准体系不统一

近年来,中国已制定了一些海洋数据相关的标准规范,但相当一部分标准不一致,如国家、地方、城市之间的空间定位基准(平面和高程)不一,数据存储管理和交换标准各异等。统一的海洋数据规范与标准体系尚未建立,使得海洋数据兼容性、可比性差,利用率低,完整性和权威性难以得到保障,海洋数据用户面对的数据集和数据格式较为混乱。

收稿日期:2020-04-15;修回日期:2020-07-30
基金项目:海南省重大科技计划(ZDKJ2019003);中船集团创新计划项目(KZ2QB9923J)
作者简介:张驰,高级工程师,研究方向为海洋信息化,电子信箱:zhangchicas@163.com
引用格式:张驰, 王瑞, 程骏超, 等. 基于区块链技术的海洋数据资源共享应用设计[J]. 科技导报, 2020, 38(21): 69-74; doi: 10.3981/j.issn.1000-7857.2020.21.008

1.2 数据信息分级分类不明确

多年来,中国海洋监测数据信息一直没有进行过分级和分类,监测数据信息的级别不明、类别不清,没有形成依据明确、科学合理的分级分类体系,使海洋监测数据信息的管理以及数据信息的应用、共享和服务等无章可循,在一定程度上影响和制约了海洋监测数据信息的利用效率[1]。

1.3 数据信息发布内容不丰富

目前中国海洋监测数据公开发布的信息都为经整编、统计、分析的评价类信息,基本不涉及原始监测数据,造成发布和共享的海洋环境监测数据信息类型单一、可利用率低,信息内容几乎多年不变。

1.4 海洋数据资源系统缺乏统一的规划与整合

目前,中国海洋科学数据资源仍缺乏综合性的国家海洋信息化规划,跨部门、跨系统间的海洋科学数据信息资源相对分散,海洋数据的使用服务统一协调性差。大多数现有的海洋数据库系统仍处于原始的离散状态,系统的性能和功能难以满足海洋数据共享服务的需求,国家亟需的对海洋开发、海洋综合管理等起支撑作用的有效信息也未被充分提取使用[2-3]。

1.5 海洋数据共享服务网络平台无法满足需求

目前,中国已开通"中国海洋信息网"等部分业务中心网站以及国家海洋监测数据传输网、海洋卫星数据传输系统等网络系统,具备了一定的数据通信与传输能力,但海洋数据共享必需的快速查询检索、传输、下载等服务能力以及数据在线处理与更新能力不足。针对无偿/有偿、公开/涉密、在线/离线、浏览/下载等相结合的共享网络访问控制、信息灾难恢复等技术和手段,还需进一步提高。

1.6 海洋数据管理与服务质量体系不健全

目前,中国尚未完全形成从原始数据采集、数据传输、数据处理、数据保管到数据应用与共享服务的海洋数据质量管理体系,针对海洋数据本身的质量评估体系建设也有待加强,使得海洋科学数据获取、处理、管理和共享各环节的准确性和可靠性难以有效管控。

1.7 海洋科学数据资源管理与共享体制亟待完善

中国在相关资料领域已从国家高度制定了一些法律法规,如《中华人民共和国测绘成果管理规定》和《地质资料管理条例》等。中国涉海部委和沿海地区也纷纷基于自身需求制定了相关规章制度,其中不乏对海洋数据的汇交、使用和共享的规定,如《海洋观测预报管理条例》规定了海洋观测资料的汇交使用,《海洋资料申请使用审批管理暂行办法》对海洋局内海洋数据的使用服务做出了规定,但由于缺少海洋资料管理的相关法律制度和高层次的海洋信息管理体制,对于海洋资料的所有权、采集权、资料的归属和转移尚未有明确规定,使盲目的海洋信息垄断现象得以蔓延,导致资源浪费,已有的信息资源不能充分利用,低水平的重复调查和研究现象严重,直接制约了中国海洋科学研究的发展。

2 基于区块链技术的海洋数据资源共享应用模式

针对涉海领域"数据孤岛"的本质原因,以区块链技术的特点与优势为基础,结合区块链技术在数据存储与资产管理方面的成功案例,进一步延伸应用于海洋数据开放共享与增值流通领域,分析技术可行性与商业可信性。

拟采用全新的去中心化系统架构与计算范式设计海洋数据资源共享应用模式,基于区块链技术,针对现存问题形成突破,实现海洋数据的自由发布、自主发现、灵活交付,交易过程安全可控、全面监管,为海洋大数据安全共享与交易提供坚实的保障。

2.1 基于联盟链概念的总体设计

联盟链是一种特殊的区块链,将区块链上的节点人为地分为若干联盟并建立联盟代理节点,由联盟代理节点执行共识算法,减少网络开销。以涉海管理部门、科研单位、相关企业为目标用户和网络节点,以用户节点为单位构建海洋数据联盟链,根据不同机构的规模大小和数据处理能力,建立强代理和弱代理两种管理节点用于执行共识算法,进而形成强联盟和弱联盟[4]。

如图1所示,联盟链中的基本节点,由联盟中

的普通用户实际控制,完成数据的发布、请求和交易功能,用户节点本身不存储数据且不记账,从而使得用户节点更加轻量级,对于网络要求较低,并能适应多源异构的环境。代理节点为联盟链中的监管和记账节点,每个联盟拥有至少一个代理节点,由联盟管理员实际控制,当联盟链中的一笔交易发生后,所有代理节点共同为这笔交易记账,因此每个代理节点都持有整个联盟链账单的一份副本,从而保证了账单的公开性和不可篡改性,满足了安全和监管的需要。此外,代理节点又分为强代理节点和弱代理节点,当一名用户请求上传数据时,将向区块链中所有强代理节点广播发布数据信息,收到该信息的全体强代理节点将首先通过该信息决定是否参加本次存储竞争,然后所有参与竞争的强代理节点共同竞争存储数据权,存储完成后生成数据索引条目,广播给区块链中所有强代理和弱代理。成功存储数据的强代理节点所在联盟可获得一定的数据上传下载带宽奖励,适合数据需求量大且服务器数据处理能力较强的强联盟。而弱代理节点则只是实时更新数据索引,并不参与数据存储,适合数据需求量较小,服务器数据处理能力较弱的弱联盟。

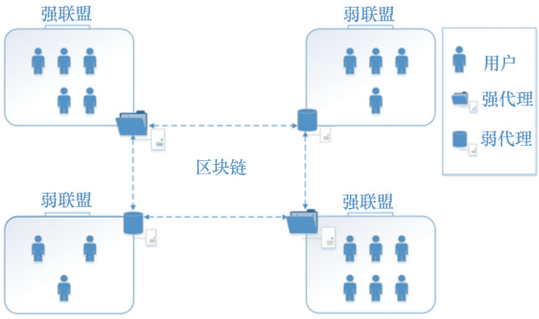

图1　海洋数据资源共享应用模式

2.2　基于改良DPOS的共识机制

共识协议或共识平台是分布式账本技术的核心。决策权越分散,系统达成共识的效率越低,但系统稳定、满意度高;决策权越集中,系统更易达成共识名单同时更易出现独裁[5]。因此,采用何种共识机制决定了整个系统的效率和稳定性。共识机制要解决的核心问题是在网络中有节点作恶时如何能够达成共识,即网络是否具有拜占庭容错能力[6]。目前,考虑到拜占庭容错问题,广泛应用的共识机制包括PoW、PoS、DPoS、PBFT等。将传统DPoS(股权授权证明)共识机制加以改良,将传统DPoS通过投票决定代理记账节点的形式改为直接将记账权分配给各联盟的代理节点,对交易进行验证和记账,从而使得每个联盟都包含一个记账节点,且记账节点为联盟管理员实际控制下的代理节点。采用DPoS共识机制很大程度上降低了区块链的网络负载,同比特币采用的工作量证明(POW)共识机制相比节省了全网算力,且依旧能够允许拜占庭容错;而舍弃授权投票过程这一改动使得整个共识过程得以简化,将记账权固定授予联盟管理员监管下的、具有更高硬件性能的代理节点,使得系统更加可靠,可监管性更强。

2.3　基于SHA256哈希算法的数据校验设计

哈希算法将任意长度的输入值映射为较短的固定长度的二进制值。SHA256算法是哈希算法的一种,将任意长度的输入映射为256位的固定长度输出,这个二进制值称为哈希值[7]。在区块链中,所有交易数据经过两次SHA256哈希运算存于Merkle树中,同时生成该次交易的数字签名,实现快速归纳和校验区块中交易的完整性和存在性[8](图2、图3)。

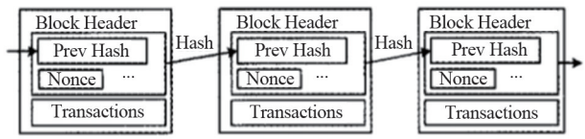

图2　区块链示意

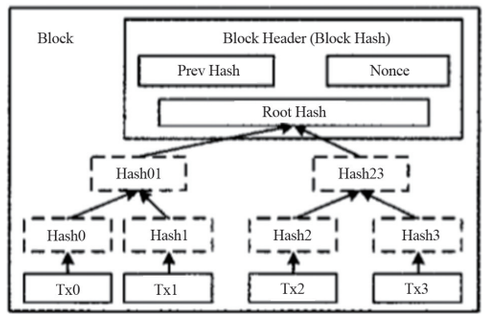

图3　Merkle树中的交易哈希

2.4 去中心化的数据存储

传统的数据存储是依赖于数据中心的中心化存储,这样的存储方式下,若数据服务器载荷过大,或遭遇黑客攻击,效率和安全性则无法保证。相比之下,去中心化存储把数据分布到多个网络节点,类似于区块链的分布式账本技术,利用了异地分布的区域性或全球性的特点,很大程度上解决了这一问题[9]。用户持有的海洋数据资源采用去中心化的方式存储,所有强代理节点共同完成海洋数据的存储工作,降低了系统的网络载荷并具备了一定的容灾能力,使整个系统更加安全、高效[10-11]。

2.5 区块链技术优势

近十年来,众多机构基于云计算等技术建设了数据中心,初步实现了信息资源共享。但在这种中心化模式下,数据中心承担了全部工作,存在诸多问题和风险。而“区块链”+信息资源共享采用去中心化模式,具有以下显著优势。

1)区块链以其可信任性、安全性和不可篡改性,让更多数据被解放出来,推进数据的海量增长。可信任性体现在区块链技术可以通过时间戳、哈希算法对海洋数据进行确权,证明一条监测数据、分析结果等存在性、真实性和唯一性。安全性体现在写入数据和读取数据的安全性以及分布式拒绝服务攻击抵抗。不可篡改性体现在区块链中分布式账本具有防篡改特性,能有效防止海洋物联网中任何单节点设备被恶意攻击和控制后带来的信息泄露和恶意操控风险。

2)区块链的可追溯特性使得数据从采集、交易、流通及计算分析的每一步记录都可以留存在区块链上,使得数据的质量获得前所未有的强信任背书,也保证了数据分析结果的正确性和数据挖掘的效果。区块链能够进一步规范数据的使用,精细化授权范围。脱敏后的数据交易流通,则有利于突破信息孤岛,建立数据横向流通机制。

3 海洋数据资源共享平台设计

因海洋数据信息量较大,直接存储于区块链上会使得网络负载过大,故只将交易信息和数据索引存于块上,而数据则交由各强代理节点进行分布式存储。拟将海洋数据联盟区块链从结构上分为数据存储链和数据交易链两部分。从功能上来看,数据存储链用于实现数据的发布和检索功能;数据交易链则用于实现节点数据的交易功能和交易查询功能。另设系统维护模块负责维护联盟链中的节点和数据,保持系统可用且高效。

在数据发布功能中,数据拥有方用户节点首先为数据定级,规定本次上传的数据为公开数据,内部数据或秘密数据。后向全网中所有强代理节点广播发布数据信息,所有强代理节点接收到此信息后根据联盟管理员设立的规则选择是否参与本次竞争,后所有参与竞争的强代理节点通过随机数和“最近最少分配”结合的动态优先级法分配数据存储权。随机数和“最近最少分配”结合的动态优先级法将数据分配权交给最近的最少分配数据的节点,这种方法的优势在于可提高分配数据的节点比例,最大程度地释放节点存储数据的潜力。获得存储权的强代理节点采用SHA256算法加密存储数据后,生成一份包含数据拥有者、存储地址和数据哈希值等信息的索引条目,向整个数据存储链上所有节点广播,共同记录该条目,将该索引条目作为区块信息存于数据存储链末端并打上时间戳,当多个代理节点确认该条目后宣告数据存储成功。

当数据请求方用户节点需要某项海洋数据时,通过数据检索功能,该节点可访问其所在联盟的代理节点,根据该用户的访问权限对代理节点维护的数据索引目录进行检索,从而得到所需数据的索引信息(图4)。

当数据请求方用户节点得到所需数据的索引信息后,数据请求方用户节点可请求交易该数据。首先,根据索引信息,请求方向数据存储方发送交易请求,存储方接收后随即向数据拥有方发送交易鉴权申请,数据拥有方回复交易鉴权确认后,数据存储方将交易数据通过数据请求方公钥加密后进行发送,请求方收到交易数据后根据索引信息中的哈希值解密数据,并向数据存储方发送交易确认信息。数据存储方收到交易确认信息后向所有代理节点广播本次交易信息,所有代理节点共同记账,

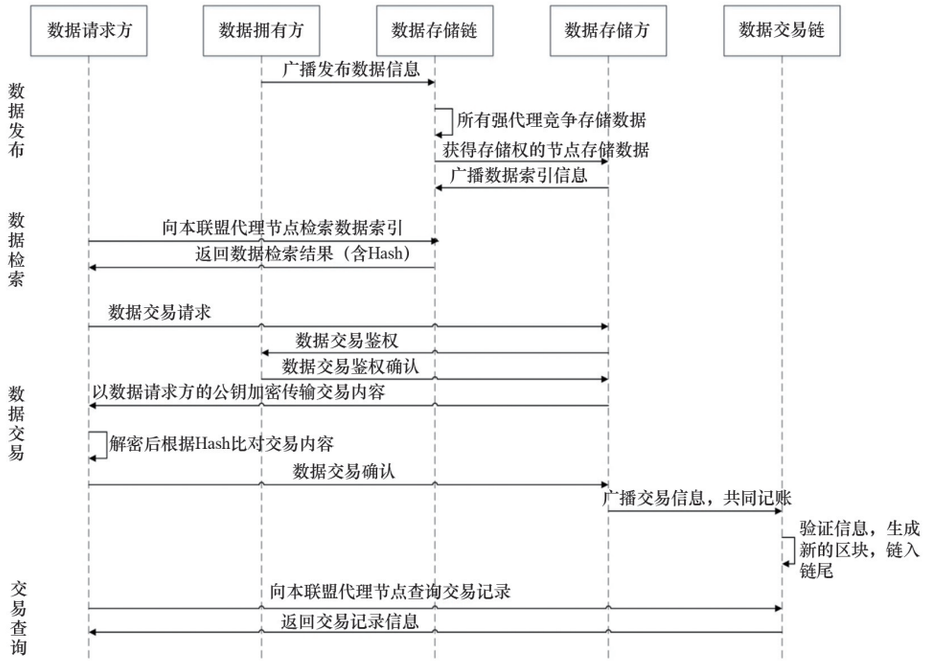

图4　基于区块链技术的海洋数据资源共享应用流程

将账目信息作为区块记录存储于数据交易链中,链至数据交易链末端并打上时间戳,当多个节点确认此单交易后宣告交易成功。因所有账目数据公开存储于区块链上,用户节点可直接向其所在联盟的代理节点根据自身访问权限对于过往账目进行查询。另外,在实际应用中,时常会出现节点加入和退出联盟的情况,为此引入维护模块。由于区块链技术采用对等网络,网内节点维护着一个在启动时可以连接的对等节点列表,当一个节点第一次启动时将自举到对等网络,从而实现与网络中所有对等节点的连接。新节点只需连接到所需加入联盟的代理节点,继而建立对等连接后再与代理节点断开,从而完成节点的加入过程。同时,对等网络的发现机制会定时监测节点是否活跃,若某一节点一段时间内未向其他节点发送过任何信息,网络会认为此节点已经断开,并将此节点信息广播至数据存储链上所有节点,若此节点曾经上传过数据,则会将该数据转至该用户所在联盟代理节点名下,由联盟管理员进行处理,从而避免大量无效数据占据存储空间[12-15]。

整个平台的设计合理,能够基于区块链技术实现海洋数据资源共享的整个过程。发布数据时,数据拥有方用户节点规定上传数据的安全等级,确保安全性;一份包含数据拥有者、存储地址和数据哈希值等信息的索引条目,会被广播至存储链上的所有节点,从而保证发布数据的可信任性;交易数据时,经过鉴权确认和数据加密处理,交易过程安全且不可篡改;同时,会将交易信息广播至所有代理节点,交易过程可信任;维护模块的引入,可以加快新节点加入联盟的过程,同时避免大量无效数据占据存储空间,保证了平台运行的效率。

4　结论

建立了由全部用户和代理节点组成的海洋数据交易区块链,将所有交易信息作为区块记录存储于交易链上,全部代理节点又组成了海洋数据存储区块链作为辅助,将数据的条目信息作为区块数据存于存储链上,数据本身则通过去中心化存储的方式存储在所有强代理节点上。

通过采用区块链技术,将海洋数据变成受保护的虚拟资产,每笔交易和数据都有确权证书,从而

保障数据所有者权益,降低数据交易成本,重塑数据市场的流通规则,激发数据交易的积极性,有利于突破信息孤岛,建立数据横向流通机制。

整体以联盟链的系统架构实现了基于区块链技术的海洋数据资源共享应用模式,实现了去中心化的系统目标,针对"海洋数据孤岛"问题给出了一种切实可行的解决方案。

参考文献(References)

[1] 洪阳, 侯雪燕. 海洋大数据平台建设及应用[J]. 卫星应用, 2016(6): 26-30.

[2] 段九如. 整合资源 推进"智慧海洋"战略[N]. 中国船舶报,2015-11-13(002).

[3] 张弘弢. 发展智慧海洋 建设海洋强国[N]. 中国船舶报, 2015-06-12(001).

[4] Dorri A, Steger M, Kanhere S S, et al. BlockChain: A distributed solution to automotive security and privacy[J]. IEEE Communications Magazine, 2017, 55(12): 119-125.

[5] Lamport L, Shostak R, Pease M. The Byzantine generals problem[J]. ACM Transactions on Programming Languages and Systems (TOPLAS), 1982, 4(3): 382-401.

[6] Castro M, Liskov B. Practical Byzantine fault tolerance [C]//Proceedings of the Third Symposium on Operating Systems Design and Implementation. New Orleans: OSDI, 1999, 99: 173-186.

[7] National Institute of Standards and Technology. 180-2: Secure hash standard (SHS)[S]. Gaithersburg: National Institute of Standards and Technology, 2012.

[8] Courtois N T, Grajek M, Naik R. Optimizing SHA256 in bitcoin mining[M]//Cryptography and Security Systems. Berlin Heidelberg: Springer, 2014: 131-144.

[9] 贾亚茹, 刘向阳, 刘胜利. 去中心化的安全分布式存储系统[J]. 计算机工程, 2012, 38(3): 126-129.

[10] 徐非, 杨广文, 鞠大鹏. 基于Peer-to-Peer的分布式存储系统的设计[J]. 软件学报, 2004, 15(2): 268-277.

[11] Haeberlen A, Mislove A, Druschel P. Glacier: highly durable, decentralized storage despite massive correlated failures[C]//Proceedings of the 2nd Conference on Symposium on Networked Systems Design & Implementation. New York: ACM, 2005: 143-158.

[12] 吴振铨, 梁宇辉, 康嘉文, 等. 基于联盟区块链的智能电网数据安全存储与共享系统[J]. 计算机应用, 2017, 37(10): 2742-2747.

[13] 杨茂江. 基于密码和区块链技术的数据交易平台设计[J]. 信息通信技术, 2016(4): 24-31.

[14] 陈何清. 基于区块链的IMIX传输系统的设计与实现[D]. 南京: 南京大学, 2016.

[15] 薛腾飞, 傅群超, 王枞, 等. 基于区块链的医疗数据共享模型研究[J]. 自动化学报, 2017, 43(9): 1555-1562.

Application design of marine data resource sharing based on block chain

ZHANG Chi[1], WANG Rui[2], CHENG Junchao[1], HE Yuan'an[1], ZHEN Jun[1]

1. Systems Engineering Research Institute, China State Shipbuilding Corporation Limited; Marine Safety System Innovation Center, Science and Technology for National Defense System Engineering Innovation Center, Beijing 100036, China

2. Institute of Software, Chinese Academy of Sciences, Beijing 100089, China

Abstract Marine data, as one of the important tools of ocean exploitation and utilization, scatter in various departments in our country, such as marine departments, scientific research institutions and other relevant enterprises. These data cannot be shared and value-added effectively, thus marine data islands are formed. This paper analyzes the fundamental reason for the marine data islands problem, then proposes a kind of marine data resource sharing application model based on block chain technology, in which the consensus mechanism of block nodes as well as the storage mode of marine data on the block chain are improved. Eventually, a marine data resource sharing platform is built to reach the goal of data sharing and value-added distribution.

Keywords ocean data; blockchain; data sharing; value-added circulation

(责任编辑 刘志远)