基于联盟区块链及 IPFS 技术的数字档案分布式应用平台设计研究*

云 健[1],王 振[1],王春霞[2]
(1.大连民族大学 计算机科学与工程学院,辽宁 大连 116650;2.大连民族大学档案馆,辽宁 大连 116650)

摘 要:在互联网时代背景下,数字档案建设正快速推进。但是,目前已有的数字档案应用系统在存储容量、分布式容灾备份、有序共享及可溯源、确保安全性等方面存在明显瓶颈。如何利用包括区块链技术和分布式技术在内的新一代信息技术研发新的数字档案分布式应用平台已成为研究热点。首先,对数字档案分布式应用平台进行了需求分析。其次,进行了数字档案分布式应用平台所需的区块链分布式存储模型及数字档案移交模型等关键模型设计、区块链账本设计和分布式存储集群设计。接着,设计了基于 Fabric 联盟链和 IPFS 技术的数字档案分布式应用平台,并给出了档案移交及其密钥生成的智能合约算法实现等环节。最后进行了系统测试。相关工作为数字档案的海量分布式存储及容灾备份、有序共享及可溯源提供了新的、更可靠的技术保障。

关键词:数字档案;联盟区块链;分布式存储;Fabric;IPFS;智能合约
中图分类号:G273 **文献标识码**:A **DOI**:10.19358/j.issn.2097-1788.2022.06.013
引用格式:云健,王振,王春霞. 基于联盟区块链及 IPFS 技术的数字档案分布式应用平台设计研究[J].网络安全与数据治理,2022,41(6):90-101.

Research on the construction of digital archives distributed application platform based on consortium blockchain and IPFS technology

Yun Jian[1], Wang Zhen[1], Wang Chunxia[2]
(1.School of Computer Science and Engineering, Dalian Minzu University, Dalian 116650, China;
2.Archives of Dalian Minzu University, Dalian 116650, China)

Abstract: In the context of the Internet era, the construction of digital archives is advancing rapidly. However, the existing digital archive application systems have obvious bottlenecks in terms of storage capacity, distributed disaster recovery backup, orderly sharing and traceability, and ensuring security. How to use the new generation of information technology including blockchain technology and distributed technology to build a new distributed application platform for digital archives has become a research hotspot. Firstly, the demand analysis of the digital archives distributed application platform is carried out. Secondly, the key model design, blockchain ledger design and distributed storage cluster design, such as the blockchain distributed storage model and digital file transfer model required by the digital archives distributed application platform, are carried out. Then, a distributed application platform for digital archives based on Fabric consortium chain and IPFS technology is constructed, and the implementation of the smart contract algorithm for archive handover and key generation is described. Finally, a system test was carried out. The related work provides a new and more reliable technical guarantee for the massive distributed storage and disaster recovery backup, orderly sharing and traceability of digital files.
Key words: digital archives; consortium blockchain; distributed storage; Fabric; IPFS; smart contract

*基金项目:2022 年大连市科协科技创新智库"大连市数字政府建设与数字经济发展统筹推进方略研究"(202211);2021 年大连民族大学研究生教育教学改革项目(202105);2022 年大连民族大学研究生精品课程建设项目(220009)

0 引言

随着第五次科技革命的到来,信息技术正在高速地蓬勃发展,呈现出多元化、网络化、多媒体化、智能化、虚拟化的趋势[1]。档案的存储方式正由纸质存储转变为数字化存储,且从过去关注档案实体管理转向高度重视知识内容管理。但是,在档案数字化的过程中,新的挑战也随之而来,如数字化档案安全性低、保密性差、有被篡改的可能,数字化档案移交接收流程较不规范、难以溯源等等。

区块链[2]技术不是单一的信息技术,它是数学、密码学、计算机科学乃至经济学、社会学的集成创新,通过“不可篡改”“全程可溯源”“全面一致性”“分布式共享”“合约智能化”“信用塑造”“正向激励”等特性来创造互联网中的信任与价值。区块链由一个个存储数据区块组成的,每个区块头都存储着各自的哈希值,同时也存储着前一个区块的哈希值,哈希值是哈希算法将任意长度的二进制值映射为固定长度的二进制值[3],它是一段数据唯一且极其紧凑的数值表示形式,具有三个安全特性,分别是碰撞阻力、隐秘性、谜题友好[4],这些特性保证了区块链难以被破解。区块链技术极大地拓展了社会协作的广度和深度,提供了互联网时代的多主体合作机制和组织形式。目前,区块链技术应用已延伸到多个社会服务领域,全球主要国家都在加快布局区块链技术发展。2019 年 10 月 24 日,中共中央政治局就区块链技术发展现状和趋势进行第十八次集体学习,会议强调要把区块链作为核心技术自主创新重要突破口,要探索利用区块链数据共享模式,实现数据跨部门、跨区域共同维护和利用,促进业务协同办理。目前,区块链技术已正式上升为国家战略。

区块链作为新一代信息技术加速突破应用,目前,包括美国、英国、瑞典、西班牙、爱沙尼亚[5-9]等国在内的全球主要国家都在加快布局区块链技术在电子政务领域的应用。具体到数字档案工作,已经受到中国政府及学界的广泛关注:一方面,2017 年国家档案局将“区块链技术在电子档案管理中的应用”纳入科技项目的立项选题指南,这标志着区块链技术正式走进中国档案界的视野;另一方面,主流学界也普遍认为区块链技术本身与数字档案工作存在契合之处,区块链技术与数字档案管理的基本要求更是高度吻合[10-11]。与此同时,也确有部分学者认为现阶段将区块链技术运用于档案管理活动不可行[10-12]。究其根本原因,有两点:一是,目前区块链在档案管理中的研究,绝大多数都是定性的理论探讨或者对“区块链+档案”特征和愿景的宏观思辨,而切实立足于“档案建设”的“区块链+档案”实际落地应用则鲜有报道;二是,鉴于我国档案工作一直遵循集中统一管理原则[13-14],以及现实工作中的数字档案的存储容量往往是巨大的,现有的区块链技术在绝对“去中心化”和“区块存储容量过小(典型应用的区块大小只有 1 MB)”两个特点上如果不加以改进,确实难以与现实的档案管理真正对接。

综合以上,本文认为对现有区块链技术及其存储模式进行科学选型、整合、改进和突破,进而切实构建起一个“区块链+档案”的实际落地应用平台具有重要意义。

1 需求分析、关键问题处理与技术选型

1.1 需求分析

(1)容灾备份的现实需求

传统的纸质档案容灾备份需求是显然的,但是数字档案容灾备份需求却往往没有引起足够重视。事实上,传统纸质档案的受灾受损过程往往是有时延的,受灾受损程度从某种角度说属于渐进过程量,往往还有救灾减损的一定余地;而数字档案的受灾受损一旦发生,比如说存放数字档案的中心化服务器在受到黑客入侵、计算机病毒植入操作系统、自然灾害使中心服务器损毁时,数字档案往往会瞬间受灾受损,甚至造成档案数据无法修复的灾难性后果。综上所述,尽管我国档案工作一直遵循集中统一管理原则,但是,这并不意味着数字档案的存储要依靠中心化服务器,恰与此相反,数字档案必须具有极强的健壮性,即在极端环境下数字档案仍然可以保证其自身的完整性,要想满足这一需求,数字档案的存储就不能依赖于中心化的服务器,而应该采取分布式存储。

(2)有序共享和权威存证的现实需求

虽然数字档案本身具有便于共享的特性,但是其不能保证共享的有序性和严谨性,究其原因是因为数字档案作为二进制电子文件在计算机网络中其内容和形式都是相对独立的,失去了固定形式,因此安全性降低、保密性变差、被篡改的可能增加、移交接收流程变得不规范且难以溯源。以上现实瓶

颈使得数字档案的原始性和法律凭证性都受到了巨大的冲击[15]。既要数字档案带来的共享便捷,又要数字档案保持其权威存证特性,这个两难问题亟需在档案界引入具有“不可篡改”“全程可溯源”“全面一致性”特征的区块链存证技术作为数字档案加以解决。区块链技术主流方向有两类,分别是传统的公有链和新一代联盟链。公有链网络无限制条件,任何想加入网络的节点都可加入,并且公有链中的记录数据任何加入网络的节点都可直接获取,具有绝对“去中心化”特征。联盟链是由指定的多个组织成员参与网络组成联盟,成员加入时,需要由证书颁发机构(Certificate Authority,CA)进行身份验证,只有身份证书和私钥文件验证通过的成员才能在网络中通信,并且联盟链网络中的数据只有具有了一定权限的成员才能调阅。本文认为需要充分利用新一代联盟区块链技术对传统区块链技术的“不可篡改”“全程可溯源”特征进行继承,对公有链技术的绝对“去中心化”倾向加以约束与限制。

(3)数字档案海量存储的现实需求

基于磁、光、电等多种载体的数字档案需要海量存储,但是经典区块链中的每一个“区块”存储容量极其有限,以区块链技术典型应用——比特币为例,其区块链网络中的每个区块大小仅有1 MB,比特币区块链网络从2008年诞生至今,完整的比特币区块链网络大小才只有400 GB,即一个TB级别的硬盘驱动器就可以覆盖这些数据。显然,以经典区块链技术中现有的存储单位和存储容量绝无可能满足数字档案海量存储需求。因此,本文认为数字档案分布式应用平台要想在有序共享、权威存证和海量存储之间取得精准平衡,就必须通过技术改进为区块链网络扩容。

1.2 解决的关键问题

(1)针对数字档案容灾备份和海量存储的问题,本文设计了一种区块链分布式存储模型,将分布式存储技术与联盟链结合,把档案文件存储在分布式网络中,使存储容量扩大,而把档案文件的属性信息存储在区块链中。与此同时,利用联盟区块链平台设计智能合约,使用联盟链引入私有数据存储方式,使某平台中的保密档案可以不必被链中所有平台共享,对不同密级的档案文件进行分类多账本存储,并且各个账本采用不同的加密方式,对不同级别的用户开放相应的档案操作权限,实现档案存储的高效性与保密性。

(2)针对目前数字档案共享无序、不规范、难溯源的问题,本文提出了一种基于分布式技术和联盟链智能合约技术的档案移交模型,通过生成私网密钥,隔离其他非联盟成员,搭建分布式存储网络,通过共享密钥的方式将各个存储组织节点相互关联,同时,在数字档案移交过程中,通过智能合约生成移交密钥,完成档案移交权限的限定。将数字档案多个存储平台组成了一个联盟网络,使得各个平台拥有一个或多个区块链节点,在保证安全的同时,可以方便快捷地进行档案的有序共享工作。

(3)以构建上述两个模型为基础,在数字档案分布式应用平台的具体构建中,解决了身份控制验证问题、档案分类存储问题、档案移交认证与调阅溯源等系统实现阶段的关键问题。

1.3 数字档案分布式应用平台的技术选型

(1)联盟区块链技术选型。Fabric[16]是Hyper ledger中的区块链平台,由Digital Asset和IBM提供,是联盟链的代表。Fabric具有多通道通信、分化Peer节点等特点,它通过Blockchain Service将区块文件存放在不同的节点中,进而降低网络的安全风险,当一个或几个节点出现问题或者遭受攻击时,其他节点不受影响。Fabric中智能合约(Smart Contract)也称为链码(Chain Code),主要包含打包合约、部署合约、同意合约、检查合约、提交合约以及查看合约。Fabric只允许链中的各个机构间传递交易数据,其他非链中机构无权访问,保证了数据的安全与可靠性。综合以上特点,本平台采用Fabric2.0联盟链作为档案存储与应用的区块链底层网络,让多个档案存储平台可以共同通信,方便档案移交接收,同时,档案数据的安全性也得到保证。

(2)分布式存储技术选型。星际文件系统(Inter Planetary File System,IPFS)是基于内容来寻址的、分布式的新型超媒体传输系统。IPFS技术诞生于2014年,由Juan Benet创建[17],Protocol Labs协议实验室发展。IPFS提供了弱冗余、高性能的集群存储方案,通过分布式哈希表(DHT)、块交换协议(BitTorrent)和自验证文件系统(SFS),创建了一个全新的分布式存储模式,它更加快速、更加安全。IPFS将存储在其系统中的每个文件都赋予一个唯一的Hash指纹,同时IPFS也会根据文件指纹追踪其历史修改记录,以上特点应用于档案存储平台,对于档案

的追踪溯源有很大的帮助。同时IPFS中包含着如Muti formats、LibP2P、IPLD等多个功能模块[18]，对于档案文件的加密水平和资源利用率等有很大的提升。综合考虑以上特点，本平台采用IPFS作为档案存储与应用的分布式存储技术。

2 关键模型功能、区块链账本和分布式存储集群设计

2.1 分布式存储模型设计

首先，进行了档案分类存储设计，具体包含两个方面：一是档案文件类型的分类存储，二是档案密级的分类存储。对于档案文件类型的分类存储，首先由用户自行选择要上传的档案文件类型，之后前端根据用户所选档案文件类型限制要上传的档案文件格式，上传之后由后端按照智能合约的相关规则将档案源文件存储到相应位置。对于档案密级(绝密、机密、一般)的分类存储，采用Fabric中隐私数据(Private Data)根据档案密级对档案文件信息进行再加密，同时采用多账本分类存储的方式，将不同密级的档案文件信息存储在不同的区块链账本中，保证档案文件信息的安全性。

其次，通过Fabric中的智能合约来设计联盟链网络成员共同认可的业务逻辑。智能合约分为用户智能合约、公共档案智能合约和私有档案智能合约三部分，用户合约主要涉及用户的登录、注册、权限分配及相关索引的构建功能，档案合约主要涉及档案数据的上传、下载、转移、共享、历史记录溯源等功能。在录入档案信息时，依据档案保密级别，控制层会调用相关智能合约，系统会根据智能合约的相关规定，对档案进行分账本存储。

(1)绝密档案：每个加入Fabric网络通道中的档案存储平台除了安装用户智能合约和公共档案智能合约外，还会安装一个特定档案智能合约，此合约会根据不同的档案存储平台自动生成一个独立的区块链账本，用于存储档案平台的绝密档案数据信息，此账本对于平台内的低权限用户以及其他档案存储平台的用户来说将无权访问。

(2)机密档案：公共智能合约会调用Get Transient()API，使得机密档案中的保密信息在提交到有关Endorsing Peer节点背书时，会依据智能合约提案(Chain code Proposal Payload)的相关规则，对其中的保密字段隐藏，在提交阶段，授权的节点将会检查策略，自己是否有权限访问私有数据(Private Data)，如果有，检查transient data store字段，观察是否在背书阶段拿到了私有数据(Private Data)。如果没有，会从其他节点去拉取。在验证和提交阶段，私有数据(Private Data)将会被存储到相关组织的私有数据(Private Data)账本中，同时把transient data store删除。没有档案拥有者的许可，网络中的其他组织无法查阅其中的保密内容。

(3)一般档案：该类档案属性信息公开，通过公共智能合约调用Put State()API，将档案数据存入普通账本中，允许档案存储平台的所有用户访问，其他档案存储平台的用户访问时也可以通过相关功能进行跨链访问。

为了确保Fabric网络通道内的各个档案存储机构能够安全地存储各个分类的档案数据，本文设计出多账本共同存储档案数据的方案模型，依据档案的不同分类——绝密、机密、一般，调用指定的智能合约，在区块链中进行分账本存储档案数据，同时将档案原件存储到IPFS网络存储集群中。具体模型设计流程如图1所示。

2.2 数字档案移交模型设计

档案的移交和接收，不仅仅是档案源文件的转移，更重要的是档案存储管理的主要责任主体的变更，针对传统数字档案存储平台在各个档案机构间转移档案数据时难统一、难溯源、不方便等问题，提出一种新的数字档案移交模型的设计方案，当存储平台的某个档案源文件或者某类档案源文件需要移交给区块链网络中其他的数字档案存储平台时，档案管理员只需执行相应的操作步骤，就能保证移交接收过程责权明确，杜绝安全隐患，从源头上保证数字档案的可靠与可信。具体设计流程如下：

(1)移交方生成移交密钥：通过智能合约确定移交方用户身份，判断当前用户是否具有权限进行档案移交操作，如果没有权限则驳回，如果有权限，则智能合约会将档案管理员传入档案移交服务的档案ID、档案Hash、档案源文件IPFS CID以及接收方MSP组合成一个新的结构体Transfer Agreement，通过Create Composite Key()算法将结构体进行加密处理并生成transfer Agree Key，存入移交方的区块链网络的私有数据集合(Private Data Collection)中，同时将要移交档案的信息传递给接收方。

(2)接收方选择是否同意移交：由接收方根据移交方提供的待移交档案的相关信息，确认是否可接

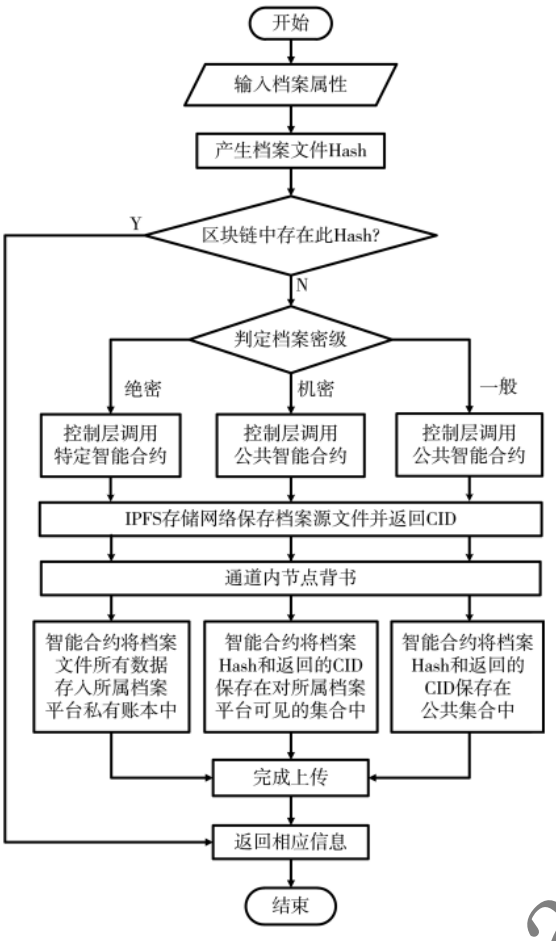

图 1 档案分布式存储模型流程图

收档案源文件。若同意移交,则智能合约会产生相同的移交密钥并存入接收方的区块链网络的私有数据集合中,移交方可进行下一步的移交操作;若不同意移交,则移交方可将移交密钥删除并选择其他档案平台接收方。

(3)移交方进行档案移交:移交方在移交档案文件时,由前端输入接收方 MSP 的名称代码,由智能合约存入暂态数据集合“archive_receiver”中,之后智能合约通过 verify Agreement( )算法调用 Fabric 的Get Private Data Hash( ) API,通过对私有数据的 Hash 比对,判定接收方是否已同意接收档案文件。在判定结果为真时,会解密存储在区块链账本中的移交密钥,获取接收方名称代码,将档案文件的所有者更改为接收方的名称代码,同时,也会保留其原所有者的历史记录,方便档案溯源。

(4)移交方删除档案信息:在档案移交完成之后,智能合约会自动将移交方私有数据集中的相关档案信息和移交密钥删除。

图 2 给出了数字档案移交模型时序图。

2.3 数字档案阅档与溯源功能设计

数字档案阅档与溯源功能主要涉及的是平台中的用户对存入区块链网络中的档案属性数据进行档案查阅和档案溯源。由于本平台采用了区块链网络作为底层通信网络,当用户选择档案历史溯源后,档案溯源服务会从前端获取到要查询的档案 ID(主键),直接通过 GetHistoryForKey( )接口从区块链账本中获取档案的历史数据,并通过 Iterator 迭代器将所有档案数据全部取出,存入 HistoryQueryResult 结构体中,通过绪论与反序列化处理将档案历史数据返回给前端。具体的档案阅档溯源时序如图 3 所示。

2.4 区块链网络账本设计

区块链中的账本数据与传统的数据库存储方式有很大差别,传统数据库一般只会保存最新的数据记录,而区块链账本会存储数据所有历史记录,区块链账本会将传入区块链的每一条数据记录都以区块的形式存储,并且区块高度与区块 Hash 一旦生成便无法改变,它也因此具有了有序和防篡改的特征。本文所使用的 Fabric 中的区块链账本由两部

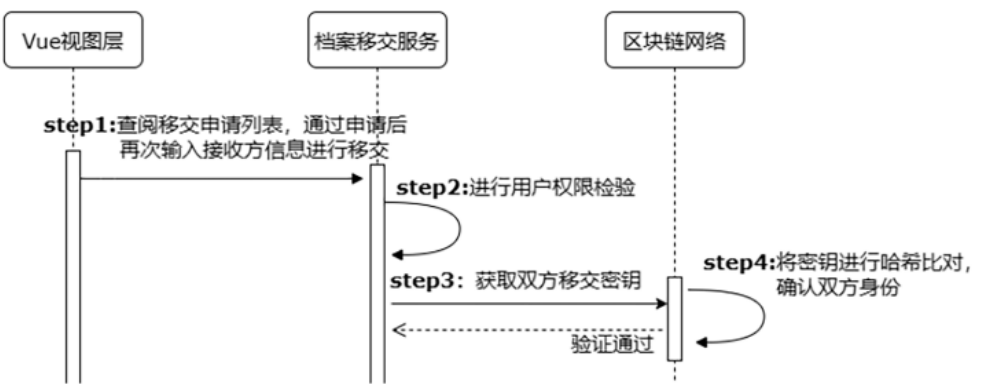

图 2 档案移交时序图

分组成，分别是世界状态(World State)和区块链(Blockchain)。在 World State 中所存储的数据的结构是以键值 key-value 的形式存储的，所存储的数据记录是最新的。世界状态中还存在着一个关键属性版本号——Version，它的作用就是记录数据的版本号，每当记录被更改，版本号就会递增。Blockchain 由一个个区块有序排列并 Hash 加密组成，主要用于记录数据的历史信息，由于 Hash 链是很难被破坏的，因此区块链具有防篡改的特性。依据本文设计的分布式存储模型，进一步将区块链网络账本设计为私有账本(Private Ledger)与公共账本(Public Ledger)两种结构共同存储档案数据的模式，如图 4 所示。

两种结构的区块链账本都包含 World State 与 Blockchain，但其中的 World State 所包含数据集合是完全不同的，私有账本(Private Ledger)中只包含私有数据集合(Private Data Collection)，用于存储档案密级为“绝密”的档案相关属性。公共账本(Public Ledger)中包含私有数据集合(Private Data Collection)

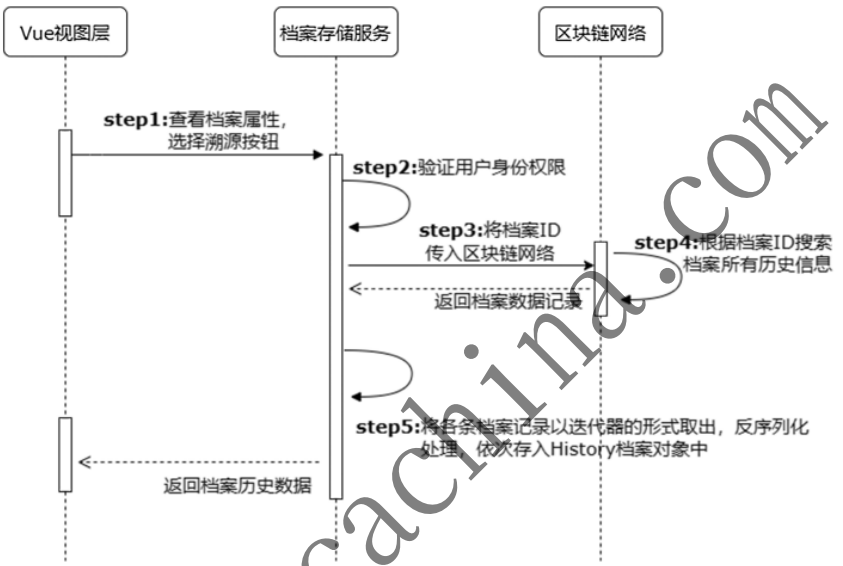

图 3 数字档案阅档与溯源时序图

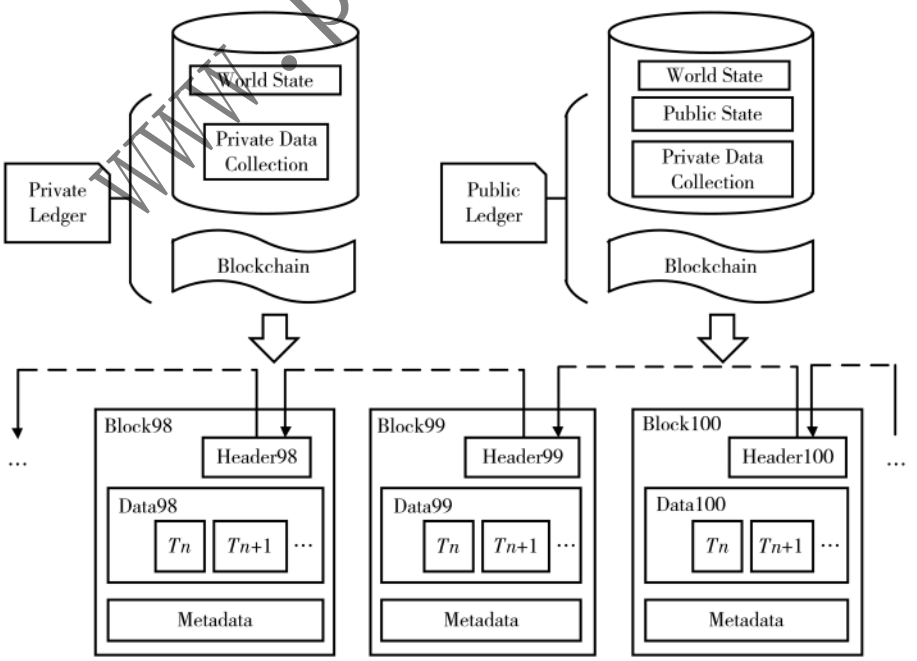

图 4 区块链网络账本结构设计

和公有数据集合(Public State)两部分,分别用于存储密级为“机密”和“一般”的档案相关属性。两个区块链账本都会将所存储的档案数据打包成区块,存入区块链网络中,便于对档案数据进行溯源。

2.5 IPFS 网络存储集群设计

身份控制验证是区块链数字档案存储平台开发的一个重要环节,本文通过 Fabric 搭建联盟链,之后,通过 IPFS 搭建存储集群形成双网络环境,其加入的组织成员身份必须得到验证才能参与网络通信,以确保区块链网络中档案数据的安全可靠。在 Fabric 搭建的联盟链网络中,由 cryptogen 工具和 FabricCA 两个组件生成各个 MSP 的 X.509 身份证书和密钥文件,实现对网络中的各个档案存储平台身份的注册与验证,在 IPFS 网络存储集群中,将采用共享同一密钥的方式确保各个组织成员的安全通信。不同组织的用户在具有一定权限后,可以访问其他档案存储节点的档案源文件信息,需要搭建一个 IPFS 私有网络存储集群,方便联盟链网络内的组织访问档案数据,同时,阻断与外部 IPFS 网络的联系,保证档案数据的安全可靠。并且,备份功能对于也是至关重要的。因此 IPFS 网络结构中包含两类节点,分别是存储节点(Store Peer)和备份节点(Backup Peer),存储节点(Store Peer)用于存储档案源文件信息以及向私有网络中共享档案数据,备份节点(Backup Peer)则用于对档案源文件的备份,具体的 IPFS 网络结构如图 5 所示。

3 平台构建与智能合约等实现

3.1 平台总体架构

根据“高内聚,低耦合”的软件工程原则,本文对平台总体功能进行了合理构建,所构建的数字档案分布式应用平台总体架构如图 6 所示,共分为四个部分,分别是前端界面、服务模块、区块链网络模块和 IPFS 网络存储集群模块。用户通过前端操作界面(Vue 和 Bootstrap 搭建)对档案文件进行相应的操作,将档案文件数据信息传给后端业务处理层(Gin 框架搭建),业务处理层将数据进行类别转换并作出相应处理,再传入区块链网络数据存储层,将数据进行相应的存取或修改,实现完整的档案数据操作流程。

3.2 平台功能划分

如图 7 所示,数字档案分布式应用平台包括了四个功能模块,分别是档案管理模块、档案存储模块、档案移交模块、系统管理模块。

档案管理模块:包括信息上传、文件下载、查阅溯源和容灾备份等子模块。信息上传主要指将档案数据的相关属性以及档案源文件存入到区块链网络数据存储账本中;文件下载主要指为档案文件提供源文件下载的功能;查阅溯源主要指对档案文件的查阅查询以及文件溯源的功能;容灾备份主要指提供对档案数据的备份还原功能。

档案存储模块:包括哈希计算、加密上链和分布存储三个子模块。哈希计算主要指对上传的档案原文件通过 SHA-256 的加密方式对其进行哈希计算,产生唯一的一个 Hash 值,用于文件验证;加密上链

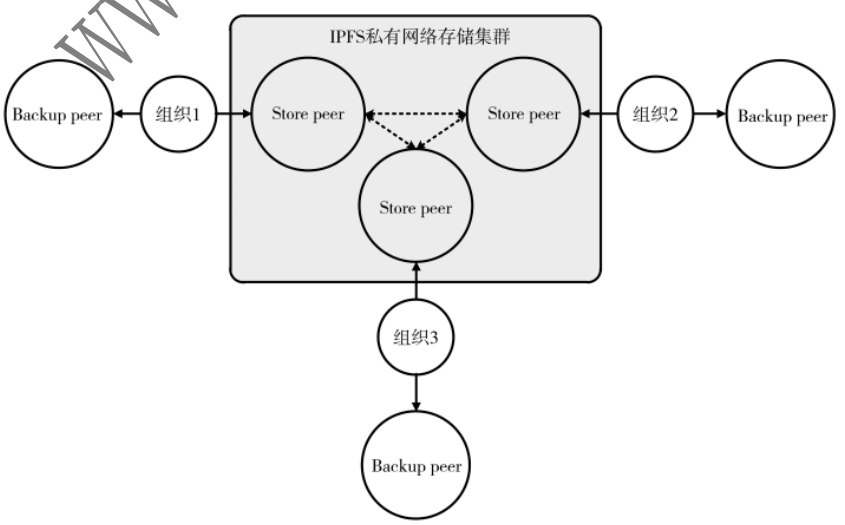

图 5 IPFS 网络存储集群设计

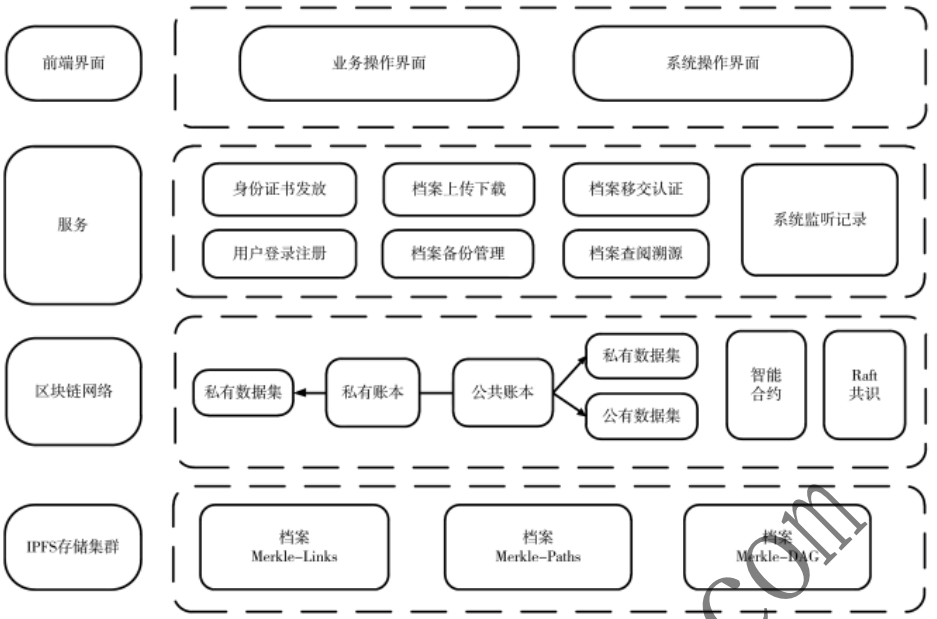

图 6 平台总体架构

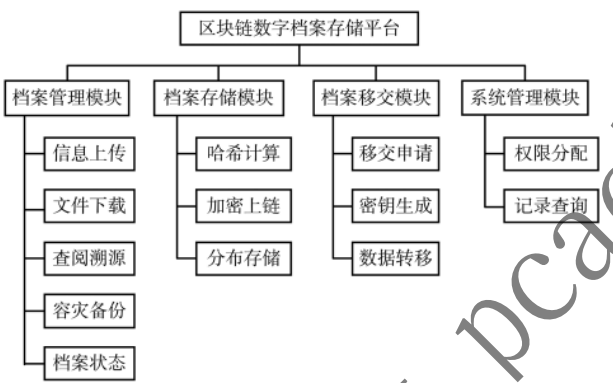

图 7 平台功能模块划分

主要指平台通过智能合约对档案属性数据以区块的形式进行加密，并存入区块链账本中；分布存储主要指对不同密级的档案属性数据进行分账本存储，同时将档案原文件存入到 IPFS 网络存储集群中。

档案移交模块：包括移交申请、密钥生成和数据转移三个子模块。移交申请主要指如果平台节点 A 想要把某个档案的相关数据移交给平台节点 B 时，需要向平台 B 进行移交申请；密钥生成主要指当平台 B 同意平台 A 的移交申请后，智能合约将为平台 A 自动生成一个移交密钥，用于移交验证；数据转移主要指平台 A 向平台 B 进行移交档案时的数据处理流程。

系统管理模块：包括权限分配和记录查询两个子模块。权限分配主要指系统管理员对平台中的其他两类用户的权限分配；记录查询主要指系统对用户的相关操作记录以及系统区块链网络的运行状态进行综合监控。

3.3 智能合约算法实现

区块链网络的功能实现核心是智能合约算法。在本平台中主要采用 fabric-contract-api-go Go 语言类库 API 对智能合约进行算法的设计与实现，本文以伪代码的形式对生成移交密钥、档案移交算法等平台核心功能的智能合约算法实现方法进行说明。

档案区块记录结构：

```
ArchiveDetails struct {
    Hash          string `json:"hash"`
    DataHash      string `json:"data_hash"`
    BlockNumber   string `json:"block_number"`
    PreviousHash  string `json:"previous_hash"`
    TxID          string `json:"tx_id"`
}
```

机密档案加密结构：

```
ArchivePrivateDetails struct {
    ObjectType    string `json:"objectType"`
                                //archivesInfo_private
    ID            string `json:"id"`
    Hash      string `json:"hash"`
    IpfsHash      string `json:"ipfs_hash"`
    IpfsBackup    string `json:"ipfs_backup"`
}
```

档案移交密钥结构：

```
TransferAgreement   struct {
    ObjectType      string `json:"objectType"`
                                //transferAgreement
    ArchiveID       string `json:"archive_id"`
    Hash            string `json:"hash"`
    ReceiverMSP     string `json:"receiver_msp"`
}
```

档案移交生成密钥算法:

```
1: procedure Procedure ParseLog
2:   ID_client ← submittingClientIdentity(ctx)
3:   transientMap ← ctx.GetStub().GetTransient()
4:   PrivateDetail_sarchive ← Json ← transientMap
     (archive agreement)
5:   Obj_archive ← GetArchive(PrivateDetail_sarchive.hash)
6:   verifyClientOrgMatchesPeerOrg(ctx)
7:   Key_transferAgree ← CreateCompositeKey(Obj_transferAgreement)
8: end procedure
```

档案移交算法:

```
1: procedure Procedure ParseLog
2:   transientMap ← ctx.GetStub().GetTransient()
3:   TransferInput_archive ← Json ← transientMap
     (archive receiver)
4:   verifyClientOrgMatchesPeerOrg(ctx)
5:   verifyAgreement(ctx, Hash, archive.Owner,
     archive.ReceiverMSP)
6:   Agreement_transfer ← GetTransferAgreement()
7:   Owner_archive ← ReceiverMSP_transferAgreement
8: end procedure
```

档案移交生成密钥算法与档案移交算法共同组成了档案移交智能合约,通过移交密钥确定移交权限,移交方档案管理员在移交档案时,通过智能合约的算法约束,将档案的所有者更改为接收者。

3.4 IPFS 相关功能实现

为了方便用户在 IPFS 网络存储集群中快速高效地进行档案源文件地上传、下载和备份等操作,本平台使用了 go-ipfs-api 类库来搭建 IPFS 网络存储集群的后台存储服务,主要包括连接网络、显示节点、档案上传、档案下载和获取档案文件状态等功能,具体使用的函数结构为:

```
type IpfsApi interface {
    GetIPFS() *shell.IdOutput
    SwarmPeers() *shell.SwarmConnInfos
    UploadIPFS(file[]byte)(string, error)
    CatIPFS(CID string)([]byte, error)
    GetArchiveStat(CIDstring)(*shell.ObjectStats, error)
}
```

4 系统测试

对所开发的基于 IPFS 和 Fabric 的数字档案分布式应用平台进行了全方位测试,囿于篇幅,本文选取档案上链存储、档案移交接收、档案修改溯源进行测试说明及测试成功页面展示。

4.1 测试环境

采用如表 1 所示的环境进行系统测试。

表 1 系统测试配置与环境

软硬件配置	环境
系统	Ubuntu 20.04.2 LTS 64-bit
CPU	Intel® Core™ i7-9750H CPU@ 2.60 GHz 2.59 GHz
内存	8 GB DDR4
Fabric-SDK-go	1.0.0
Fabric	2.3
IPFS	0.9
Docker	20.10.6
浏览器	Firefox
TAPE	0.1.2
Orderer 节点/个	1
Peer 节点/个	3
IPFS 节点/个	4

4.2 测试用例设计及实施

(1)档案上链存储测试

档案上链存储功能将只能由档案管理员操作,表 2 为档案上链存储的测试内容以及结果。图 8 为档案上链存储成功后的档案列表界面。从系统界面以及测试结果中可以看出,本平台成功地将档案属性数据上传到了区块链网络中,并且每一条档案属性数据都有对应的唯一区块高度、数据哈希以及上传时的事务 ID 信息,同时,每一条档案信息都可以进行溯源,利用区块链的不可篡改性确保了档案信息的安全。平台将档案源文件存储在了 IPFS 网络存储集群中,相比于只采用区块链存储档案源文件,大大增加了数字档案源文件的分布式存储能力。

(2)档案移交接收测试

档案的移交接收功能将只能由档案管理员操作。档案管理员可以从档案列表里选择要移交的档案,进行移交申请,待移交申请由接收方审核通过后,进行正式移交,由系统将档案所有者更改为接

收方，同时删除移交方中的对应档案信息，不需要将档案文件进行下载后再移交，方便快捷，同时由于 IPFS 私有网络集群中的文件格式都转换为了字节类型，网络中的各个组织对移交后的档案文件的格式处理也极为方便。表 3 为档案移交接收的测试内容以及结果。

(3)档案修改溯源测试

档案的修改与溯源功能将只能由档案管理员操作。档案管理员可以从档案列表里选择要修改或溯源的档案，档案被修改后将存储到新的区块中，通过智能合约中的 GetHistoryForKey( )接口，档案管理员可以对当前档案进行溯源，查看修改记录以及

表 2 档案上链存储测试用例

用例编号	testUploadArchive01			
用例名称	档案上链存储用例			
测试目标	检验平台的档案上链存储功能是否完善			
前置条件	联盟链网络和 IPFS 网络存储集群已正常启动，后端档案存储服务正常			
测试情况				
	编号	测试步骤及输入	测试结果	预期结果
1	档案管理员在档案上传页面录入相关的档案属性并在本地选择要上传的档案源文件，点击上传按钮进行上传	前端提示成功上传，并且在 CouchDB 和 IPFS 网络存储集群中能查询到档案信息	符合预期	
2	档案管理员在档案上传页面未录入档案属性或者未选择上传档案源文件	前端提示错误信息，上传失败	符合预期	
3	以普通用户或系统管理员身份登录系统，进行档案上传操作	前端未出现上传功能，上传失败	符合预期	

表 3 档案移交接收测试用例

用例编号	testTransfer_Receive01			
用例名称	档案移交接收用例			
测试目标	检验平台的档案移交接收功能是否完善			
前置条件	联盟链网络和 IPFS 网络存储集群已正常启动，后端档案移交服务正常			
测试情况				
	编号	测试步骤及输入	测试结果	预期结果
1	档案管理员在档案列表页面查看档案属性，点击移交按钮，向指定平台申请移交 接收方收到移交申请，并同意接收 档案管理员进行档案移交	档案所有者更改为接收方，同时移交方中的档案信息被删除	符合预期	
2	档案管理员在档案列表页面查看档案属性，点击移交按钮，向指定平台申请移交 接收方收到移交申请，但不同意接收	档案管理员选择删除申请，档案所有者未发生改变	符合预期	

图 8 上传成功后档案列表界面

文件信息。表 4 为档案修改溯源测试内容以及结果,图 9 为档案溯源界面。

5 结论

我国数字档案建设正在快速推进,但是,以目前已投入使用的数字档案存储或管理系统来共享、处理大数据时代背景下的海量数字档案文件,在存储容量、分布式备份、有序共享及可溯源、保安全等方面已有明显短板。本文提出并开发构建的基于 Fabric 联盟链和 IPFS 技术的数字档案分布式应用平台为数字档案的海量分布式存储及容灾备份、有序共享及可溯源提供了新的、可靠的技术保障。

参考文献

[1] 李芳,彭嘉琳.电子信息技术发展中的问题及发展趋势分析[J].科学中国人,2016(8X):203-204.

[2] 袁勇,王飞跃.区块链技术发展现状与展望[J].自动化学报,2016(42):481-494.

[3] BASHIR I.Mastering blockchain[M].Packt Publishing,2017.

[4] NARAYANAN A,BONNEAU J,FELTEN E,et al.Bitcoin and cryptocurrency technologies:a comprehensive introduction[M].Princeton University Press,2016.

[5] 刘卫铠,杨智勇.区块链在电子政务服务中的应用研究[J].档案与建设,2021(5):20-26.

[6] 石菲.世界各国政府谁最爱区块链[J].中国信息化,2018(8):16-23.

[7] 唐文剑,吕雯,林松祥,等.区块链将如何重新定义世界[M].北京:机械工业出版社,2016.

[8] 韩璇,袁勇,王飞跃.区块链安全问题:研究现状与展望[J].自动化学报,2019(1):206-225.

[9] Bitnation 与爱莎尼亚在区块链上开展政务管辖[EB/OL].(2016-10-28).http://www.bitecoin.com/

表 4 档案溯源测试用例

用例编号	testHistoryArchive01			
用例名称	档案溯源用例			
测试目标	检验平台的档案溯源功能是否完善			
前置条件	联盟链网络和 IPFS 网络存储集群已正常启动,后端档案移交服务正常			
测试情况				
	编号	测试步骤及输入	测试结果	预期结果
1	档案管理员在档案列表页面查看档案属性,点击修改按钮,修改档案信息	档案信息修改成功	符合预期	
2	档案管理员将之前未修改的档案上传	平台提示档案信息已存在,无法成功上传	符合预期	
3	档案管理员将未修改的档案进行溯源操作	平台提示档案历史记录不存在	符合预期	
4	档案管理员在档案列表页面查看档案属性,点击溯源按钮,查看档案修改历史	平台显示出档案的历史修改记录以及区块的高度和文件哈希值	符合预期	

图 9 档案溯源界面

online/2015/11/16308.html.

[10] 于欢欢,程慧平.区块链技术在国内电子档案管理中的应用研究述评[J].档案与建设,2021(5):27-33.

[11] 王子鹏,李璐璐.基于区块链技术的电子文件管理模式研究[J].浙江档案,2018(2):18-20.

[12] 李高峰,马国胜,胡国强.现阶段区块链技术在档案管理中不可行分析[J].档案管理,2018(5):30-32.

[13] 王国才.档案工作应用区块链技术的探索与实践——"'互联网+政务服务'背景下区块链技术在'广域数字档案馆体系'中的应用"课题思考[J].中国档案,2020(12):32-33.

[14] 李曲直,韩丽."区块链+人事档案"管理应用初探[J].中国档案,2020(7):73-75.

[15] 付艳.试论电子档案的利弊及电子档案的管理[J].经济研究导刊,2014(23):289-290.

[16] 冯翔.Hyperledger Fabric 关键技术与案例分析[M].北京:机械工业出版社,2018.

[17] BENET J.IPFS-content addressed,versioned,P2P file system[J].arXiv preprint arXiv:1407.3561.

[18] CONFAIS B,LEBRE A,PARREIN B.An object store service for a fog/edge computing infrastructure based on IPFS and a scale-out NAS[C]//2017 IEEE 1st International Conference on Fog and Edge Computing (ICFEC),2017:41-50.

(收稿日期:2022-11-17)

作者简介:

云健(1975-),男,博士,教授,主要研究方向:区块链技术及应用。

王振(1996-),男,硕士研究生,主要研究方向:区块链技术及应用。

王春霞(1979-),通信作者,女,硕士,工程师,主要研究方向:档案数字化。E-mail:wx_c@tom.com。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部